

SPRINGER  
REFERENCE

Nathan Ida  
Norbert Meyendorf  
*Editors*

# Handbook of Advanced Nondestructive Evaluation

---

# **Handbook of Advanced Nondestructive Evaluation**

---

Nathan Ida • Norbert Meyendorf  
Editors

# Handbook of Advanced Nondestructive Evaluation

With 949 Figures and 43 Tables



Springer

*Editors*

Nathan Ida  
Department of Electrical and Computer  
Engineering  
The University of Akron  
Akron, OH, USA

Norbert Meyendorf  
Center for Nondestructive Evaluation  
Iowa State University Center for  
Nondestructive Evaluation  
Ames, IA, USA

ISBN 978-3-319-26552-0

ISBN 978-3-319-26553-7 (eBook)

ISBN 978-3-319-30122-8 (print and electronic bundle)

<https://doi.org/10.1007/978-3-319-26553-7>

© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

---

## Preface

The Internet of Things (IoT) and the next generation of industrial production (Industry 4.0) encompass the complete networking of all industrial areas. New production techniques as, for example, 3D printing will allow efficient on-time production of low numbers of unique parts. A significant aspect is quality and maintainability of these (sometimes) unique structures and components. Nondestructive evaluation or testing (NDE or NDT) must, necessarily, follow these trends by not only adapting NDE techniques to new technologies but also introducing the capability of cyber systems into the inspection and maintenance processes and anticipating future needs. These new challenges and capabilities will also boost the application of unconventional and new NDE principles to industrial applications.

Industry 4.0 and the ability to tailor individual components to the customer's needs will significantly impact the way we provide nondestructive inspection and evaluation. NDE must be integrated into the production process by networking with processes and production steps during manufacturing. This will result in a paradigm shift in industrial quality management and NDE. Classical concepts based on comparison of multiple similar components and statistical analysis will not be applicable under these conditions, raising the impact of the human factor. Availability of specialists capable of making the right decisions based on NDE results, knowledge about the material and the components, loading conditions, and NDE modeling of NDE experiments will be essential for the future.

This new generation of NDE specialists need to have a broad overview of conventional and new, advanced NDE techniques. A considerable amount of overview literature for common NDE techniques that are usually based on standards is available worldwide. Several organizations offer NDE training classes and certification for inspectors for the standard techniques. Other new methods that may, at present, be in experimental, laboratory stages but have the potential for application for future inspection tasks can only be found in specific scientific journals and, sometimes, may not even be considered as NDE candidates. However, to solve future NDE tasks, the specialist should have a clear understanding of what is possible without going into too many details.

The present book intends to bridge this gap between conventional common sense NDE methods of the present and the past and advanced techniques that provide and guide new opportunities for inspections for the next generation of NDE. As is often

the case, many methods described here have evolved from prior experience and from pressing needs in industrial inspection. The reader will find that the offerings in the present handbook is a healthy mixture of methods that are in limited use, those that are at various stages of development and some that are envisioned for the future.

This handbook is structured on the lines of accepted NDE principles but focuses on advanced methods of measurement or data analysis. It is not the intent of the book to introduce the basics of NDE principles. The exceptional contributions that make up this handbook were made by specialists worldwide working on advanced NDE techniques. The editors are grateful that so many excellent contributions have been submitted and are happy to present this unique overview of advanced NDE techniques. The task of keeping up to date, of course, can never be completed and any attempt at doing so can only be a snapshot of present activities. In this spirit, the online version of the handbook will be updated and enlarged in the future to keep the contents up to date.

We sincerely thank all those involved in the writing, editing, and production of this work.

June 2019

Nathan Ida  
Norbert Mayendorf

---

# Contents

## Volume 1

<b>Part I Acoustic Techniques .....</b>	<b>1</b>
<b>1 Physical Basis for Ultrasonic Acoustics .....</b>	<b>3</b>
Victor Giurgiutiu and Bin Lin	
<b>2 Ultrasonic Wavefield Imaging .....</b>	<b>43</b>
Jennifer E. Michaels	
<b>3 Acoustic Microscopy .....</b>	<b>75</b>
Frank Schubert, Martin Barth, Raffael Hipp, and Bernd Köhler	
<b>4 Acoustic Emission .....</b>	<b>115</b>
Eleni Tsangouri and Dimitrios G. Aggelis	
<b>5 Guided Wave Testing .....</b>	<b>141</b>
Paul Fromme	
<b>6 Laser-Induced Surface Acoustic Waves for Material Testing .....</b>	<b>171</b>
Dieter Schneider	
<b>7 The Acousto-elastic Effect and Its Use in NDE .....</b>	<b>235</b>
Hans-Rüdiger Herzer, Michael Mathias Becker, and Eckhardt Schneider	
<b>8 Nonlinear Acoustics .....</b>	<b>251</b>
Younho Cho and Weibin Li	
<b>9 Local Acoustic Resonance Spectroscopy .....</b>	<b>271</b>
Christian U. Grosse, Anne Jüngert, and Philipp Jatzlau	
<b>10 Nonlinear Resonant Acoustic Spectroscopy .....</b>	<b>295</b>
Bart Van Damme and Koen Van Den Abeele	

<b>Part II Optical Techniques</b>	.....	<b>325</b>
<b>11 Surface Brillouin Scattering</b>	.....	327
Arthur G. Every and J. Darrell Comins		
<b>12 Interferometric Methods in NDE</b>	.....	361
Krzysztof Patorski and Maciej Trusiak		
<b>13 Shearography</b>	.....	383
Lianxiang Yang and Junrui Li		
<b>14 White Light Interferometry</b>	.....	421
Joanna Schmit and Anna Pakuła		
<b>15 Optical Coherence Tomography for NDE</b>	.....	469
Jonas Golde, Lars Kirsten, Christian Schnabel, Julia Walther, and Edmund Koch		
<b>16 Ellipsometry</b>	.....	513
Jian Chen and Qiwen Zhan		
<b>17 Raman Scattering</b>	.....	541
Rudolph M. Erasmus and J. Darrell Comins		
<b>18 Optical Fiber Methods in Nondestructive Evaluation</b>	.....	595
Wolfgang R. Habel		
<b>Part III Electromagnetic Techniques</b>	.....	<b>643</b>
<b>19 Eddy Current Testing</b>	.....	645
Zhenmao Chen, Cherdpong Jomdecha, and Shejuan Xie		
<b>20 High-Frequency Eddy Current Techniques</b>	.....	729
Susanne Hillmann, Martin H. Schulze, and Henning Heuer		
<b>21 Eddy Current Tomography</b>	.....	757
Antonello Tamburrino and Guglielmo Rubinacci		
<b>22 Motion-Induced Eddy Current Testing</b>	.....	781
Hartmut Brauer and Marek Ziolkowski		
<b>23 Low Field Methods (GMR, Hall Probes, etc.)</b>	.....	827
Vivek T. Rathod, Portia Banerjee, and Yiming Deng		
<b>24 Micromagnetic Materials Characterization</b>	.....	881
Klaus Szielasko and Ralf Tschuncky		
<b>25 Electrochemical Techniques, Impedance, and Spectroscopy</b>	.....	899
Shengxi Li, Qixin Zhou, and Hongbo Cong		

**Volume 2**

<b>Part IV Electromagnetic Microwave and Millimeter-Wave Techniques .....</b>	<b>927</b>
<b>26 Microwave and Millimeter Wave Nondestructive Testing and Evaluation .....</b>	<b>929</b>
Nathan Ida	
<b>27 Terahertz Techniques in NDE .....</b>	<b>967</b>
Joachim Jonuscheit	
<b>28 Ground Penetrating Radar .....</b>	<b>987</b>
X. Lucas Travassos and Mario Fernandez Pantoja	
<b>Part V X-Ray Techniques .....</b>	<b>1025</b>
<b>29 Processing of X-Ray Images .....</b>	<b>1027</b>
Sergei Chakhlov	
<b>30 X-Ray Phase Contrast Methods .....</b>	<b>1053</b>
Sheridan Mayo and Marco Endrizzi	
<b>31 X-Ray Tomography .....</b>	<b>1095</b>
Johann Kastner and Christoph Heinzl	
<b>32 3D X-Ray Tomography: Basics and Latest Developments .....</b>	<b>1167</b>
Theobald O. J. Fuchs and Randolph Hanke	
<b>33 Grazing Incidence X-Ray Reflectivity and Scattering .....</b>	<b>1181</b>
Brian K. Tanner	
<b>Part VI Particle Methods .....</b>	<b>1215</b>
<b>34 Neutron Radiography and Tomography .....</b>	<b>1217</b>
Wolfgang Treimer	
<b>35 Positron Annihilation .....</b>	<b>1301</b>
Luca Chiari and Masanori Fujinami	
<b>Part VII Thermal Techniques .....</b>	<b>1347</b>
<b>36 Physical Basics of Thermal Techniques of Nondestructive Evaluation .....</b>	<b>1349</b>
Vladimir P. Vavilov	
<b>37 Passive Thermography, Thermal Imaging .....</b>	<b>1371</b>
Helmut Budzier and Gerald Gerlach	

<b>38</b>	<b>Active Thermography</b>	1401
	Helmut Budzier and Gerald Gerlach	
<b>39</b>	<b>Thermal Wave Techniques</b>	1419
	Gunnar Suchaneck, Agnes Eydam, and Gerald Gerlach	
<b>40</b>	<b>Sonothermic Techniques in Nondestructive Evaluation</b>	1479
	Xiaoyan Han	
<b>41</b>	<b>Induction Thermography of Surface Defects</b>	1497
	Udo Netzelmann	
	<b>Part VIII Special Techniques for Signal and Data Analysis</b>	1523
<b>42</b>	<b>Signal Processing for NDE</b>	1525
	Masoud Vejdannik, Ali Sadr, Victor Hugo C. de Albuquerque, and João Manuel R. S. Tavares	
<b>43</b>	<b>Digital Image Correlation Techniques for NDE and SHM</b>	1545
	Christopher Nizrecki, Javad Baqersad, and Alessandro Sabato	
<b>44</b>	<b>Structural Health Monitoring</b>	1591
	Bianca Weihnacht, Uwe Lieske, Tobias Gaul, and Kilian Tschöke	
	<b>Index</b>	1609

---

## About the Editors



**Nathan Ida** is currently distinguished professor of Electrical and Computer Engineering at The University of Akron in Akron, Ohio, where he has been since 1985. His current research interests are in the areas of electromagnetic nondestructive testing and evaluation of materials at low and microwave frequencies with particular emphasis on theoretical issues, on all aspects of modeling and simulation and on related issues stemming from research in NDE. Starting with modeling of eddy current and remote field phenomena, and continuing with high frequency methods for microwave NDE, his work now encompasses the broad aspects of computational electromagnetics where he has contributed to both the understanding of the interaction of electromagnetic fields with materials and the development of new methods and tools for numerical modeling and simulation for, and beyond, NDE. Other areas of current interest include electromagnetic wave propagation, theoretical issues in computation, as well as in communications and sensing, especially in low power remote control and wireless sensing. Much of this work has found its way into practice through industrial relations and consulting across industries as diverse as power generation, polymers, steel, medical, and software, spanning the globe. Dr. Ida has published extensively on electromagnetic field computation, parallel and vector algorithms and computation, nondestructive testing of materials, surface impedance boundary conditions, sensing, and others, in over 400 publications. He has written nine books: two on computation of electromagnetic fields (one in its second edition), one on modeling for nondestructive testing, one on nondestructive testing with microwaves, a textbook on engineering

electromagnetics (now in its fourth edition), a textbook on sensing and actuation (now in its second edition), a book on the use of surface impedance boundary conditions, and others, including on ground penetrating radar and industrial sensing based on microwaves. Dr. Ida is a life fellow of the Institute of Electric and Electronics Engineers (IEEE), a fellow of the American Society of Nondestructive Testing (ASNT), a fellow of the Applied Computational Electromagnetics Society (ACES), and a fellow of the Institute of Electronics and Technology (IET). Dr. Ida teaches electromagnetics, antenna theory, electromagnetic compatibility, sensing, and actuation, as well as computational methods and algorithms.

Dr. Ida received his B.Sc. in 1977 and M.S.E.E. in 1979 from the Ben-Gurion University in Israel, and his Ph.D. from Colorado State University in 1983.



**Norbert Meyendorf** retired in fall 2018 as deputy director of the Center for Nondestructive Evaluation and professor in the Aerospace Engineering department at the Iowa State University in Ames, Iowa. Before joining ISU in 2016 he had several appointment and ranks. The most recent are:

Branch Director at the Fraunhofer Institute for Non-destructive Testing IZFP and later IKTS, director of the International University of Dayton/Fraunhofer Research Center at the School of Engineering at the University of Dayton, organizing collaborative projects between Fraunhofer and University of Dayton, and Program Director of the Master program “Nondestructive Testing, M. Sc. (NDT)” at the Dresden International University (DIU) between 2011 and 2015.

Norbert Meyendorf continues to be active as adjunct professor for micro- and nano-NDE at the University of Dresden and adjunct professor at the Department for Chemical and Materials Engineering, University of Dayton.

He is the author or coauthor of more than 300 peer-reviewed journal articles, contributions to edited proceedings, technical reports, and numerous oral presentations on conferences, meetings, workshops, etc. He is editor in chief of the *Journal of Nondestructive Evaluation* and edited several books and conference proceedings.

His areas of expertise include solid state physics and physical analytics, welding metallurgy, materials testing, nondestructive evaluation (NDE), and structural health monitoring (SHM), for instance.

Since 2001, he has been chairman or co-chairman of several conferences within the SPIE International Symposium on Nondestructive Evaluation for Health Monitoring and Diagnostics and later the Symposium for Smart Structures and NDE. In 2005, 2006, 2012, and 2013, he was chair or co-chair of the whole SPIE Symposium. In 2018 he became fellow of SPIE.

Norbert Meyendorf was founder and chair of two expert committees of the German Society for Non-Destructive Testing (DGZfP), the Expert Committees for “Structural Health Monitoring” and “Materials Diagnostics.” Between 2016 and 2018, he reorganized and directed the ASNT Section Iowa.

---

## Contributors

**Koen Van Den Abeele** Physics, Kulak Kortrijk Campus, Kortrijk, Belgium

**Dimitrios G. Aggelis** Department of Mechanics of Materials and Constructions (MeMC), Vrije Universiteit Brussel (VUB), Brussels, Belgium

**Portia Banerjee** SGT Inc., NASA Ames Research Center, Mountain View, CA, USA

**Javad Baqersad** Department of Mechanical Engineering, Kettering University, Flint, MI, USA

**Martin Barth** Ultrasonic Sensors and Methods, Fraunhofer IKTS, Dresden, Germany

**Michael Mathias Becker** Fraunhofer Institute for Nondestructive Testing IZFP, Saarbrücken, Germany

**Hartmut Brauer** Technische Universitaet Ilmenau, Ilmenau, Germany

**Helmut Budzier** Institute of Solid State Electronics, Technische Universität Dresden, Dresden, Germany

**Sergei Chakhlov** School of Non-Destructive Testing and Security, Tomsk Polytechnic University, Tomsk, Russia

**Jian Chen** School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China

**Zhenmao Chen** State Key Laboratory for Strength and Vibration of Mechanical Structures Shaanxi ERC for NDT and Structural Integrity Evaluation School of Aerospace, Xi'an Jiaotong University, Xian, China

**Luca Chiari** Department of Applied Chemistry and Biotechnology, Graduate School of Engineering, Chiba University, Chiba, Japan

**Younho Cho** School of Mechanical Engineering, Pusan National University, Busan, South Korea

**J. Darrell Comins** School of Physics, Materials Physics Research Institute and DST-NRF Centre of Excellence in Strong Materials, University of the Witwatersrand, Johannesburg, South Africa

**Hongbo Cong** Department of Chemical and Biomolecular Engineering, Corrosion Engineering Program, The University of Akron, Akron, OH, USA

**Victor Hugo C. de Albuquerque** Programa de Pós Graduação em Informática Aplicada, Universidade de Fortaleza (UNIFOR), Fortaleza, Ceará, Brazil

**Yiming Deng** NDE Laboratory, Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA

**Marco Endrizzi** University College London, London, UK

**Rudolph M. Erasmus** School of Physics, Materials Physics Research Institute and DST-NRF Centre of Excellence in Strong Materials, University of the Witwatersrand, Johannesburg, South Africa

Microscopy and Microanalysis Unit, University of the Witwatersrand, Johannesburg, South Africa

**Arthur G. Every** School of Physics, University of the Witwatersrand, Johannesburg, South Africa

**Agnes Eydam** Institute of Solid State Electronics, Technische Universität Dresden, Dresden, Germany

**Paul Fromme** Department of Mechanical Engineering, UCL, London, UK

**Theobald O. J. Fuchs** Fraunhofer EZRT, Fürth, Germany

**Masanori Fujinami** Department of Applied Chemistry and Biotechnology, Graduate School of Engineering, Chiba University, Chiba, Japan

**Tobias Gaul** Fraunhofer Institute for Ceramic Technologies and Systems IKTS, Dresden, Germany

**Gerald Gerlach** Institute of Solid State Electronics, Technische Universität Dresden, Dresden, Germany

**Victor Giurgiutiu** University of South Carolina, Columbia, SC, USA

**Jonas Golde** Clinical Sensoring and Monitoring, Anesthesiology and Intensive Care Medicine, Faculty of Medicine Carl Gustav Carus, TU Dresden, Dresden, Germany

**Christian U. Grosse** Center for Building Materials, Chair of Non-destructive Testing, Technical University of Munich, München, Germany

**Wolfgang R. Habel** Department of Non-Destructive Testing, Division of Fibre Optic Sensors, BAM Federal Institute for Materials Research and Testing (formerly), Berlin, Germany

**Xiaoyan Han** Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI, USA

**Randolf Hanke** Fraunhofer IZFP, Saarbrucken, Germany

**Christoph Heinzl** University of Applied Sciences Upper Austria, Wels, Austria

**Hans-Rüdiger Herzer** Fraunhofer Institute for Nondestructive Testing IZFP, Saarbrücken, Germany

**Henning Heuer** Systems for Testing and Analysis, Fraunhofer IKTS, Dresden, Germany

**Susanne Hillmann** Accredited Test Lab, Fraunhofer IKTS, Dresden, Germany

**Raffael Hipp** Ultrasonic Sensors and Methods, Fraunhofer IKTS, Dresden, Germany

**Nathan Ida** Department of Electrical and Computer Engineering, The University of Akron, Akron, OH, USA

**Philipp Jatzlau** Center for Building Materials, Non-destructive Testing, Technical University of Munich, München, Germany

**Cherdpong Jomdecha** State Key Laboratory for Strength and Vibration of Mechanical Structures Shaanxi ERC for NDT and Structural Integrity Evaluation School of Aerospace, Xi'an Jiaotong University, Xian, China

**Joachim Jonuscheit** Fraunhofer Institute for Industrial Mathematics ITWM, Kaiserslautern, Germany

**Anne Jüngert** University of Stuttgart, Stuttgart, Germany

**Johann Kastner** University of Applied Sciences Upper Austria, Wels, Austria

**Lars Kirsten** Clinical Sensoring and Monitoring, Anesthesiology and Intensive Care Medicine, Faculty of Medicine Carl Gustav Carus, TU Dresden, Dresden, Germany

**Edmund Koch** Clinical Sensoring and Monitoring, Anesthesiology and Intensive Care Medicine, Faculty of Medicine Carl Gustav Carus, TU Dresden, Dresden, Germany

**Bernd Köhler** Accredited NDT Test Lab, Fraunhofer IKTS, Dresden, Germany

**Junrui Li** Optical Laboratory, Department of Mechanical Engineering, Oakland University, Rochester, MI, USA

**Shengxi Li** Department of Chemical and Biomolecular Engineering, Corrosion Engineering Program, The University of Akron, Akron, OH, USA

**Weibin Li** School of Aerospace Engineering, Xiamen University, Xiamen, China

**Uwe Lieske** Fraunhofer Institute for Ceramic Technologies and Systems IKTS, Dresden, Germany

**Bin Lin** Department of Mechanical Engineering, University of South Carolina, Columbia, SC, USA

**X. Lucas Travassos** Federal University of Santa Catarina, Joinville/Florianopolis, Brazil

**Sheridan Mayo** CSIRO Manufacturing, Clayton, VIC, Australia

**Jennifer E. Michaels** School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA

**Udo Netzelmann** Fraunhofer-Institute for Nondestructive Testing IZFP, Saarbrücken, Germany

**Christopher Nieszrecki** Department of Mechanical Engineering, University of Massachusetts Lowell, Lowell, MA, USA

**Anna Pakuła** Warsaw University of Technology, Warsaw, Poland

**Mario Fernandez Pantoja** Universidad de Granada, Granada, Spain

**Krzysztof Patorski** Department of Mechatronics, Institute of Micromechanics and Photonics, Warsaw University of Technology, Warsaw, Poland

**Vivek T. Rathod** NDE Laboratory, Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA

**Guglielmo Rubinacci** Department of Electrical Engineering and Information Technology, Università degli Studi di Napoli Federico II, Napoli, Italy

**Alessandro Sabato** Department of Mechanical Engineering, University of Massachusetts Lowell, Lowell, MA, USA

**Ali Sadr** School of Electrical Engineering, Iran University of Science and Technology (IUST), Narmak, Tehran, Iran

**Joanna Schmit** 4D Technology, Tucson, AZ, USA

**Christian Schnabel** Clinical Sensing and Monitoring, Anesthesiology and Intensive Care Medicine, Faculty of Medicine Carl Gustav Carus, TU Dresden, Dresden, Germany

**Dieter Schneider** Fraunhofer Institute for Material and Beam Technology (IWS), Dresden, Germany

**Eckhardt Schneider** Fraunhofer Institute for Nondestructive Testing IZFP, Saarbrücken, Germany

**Frank Schubert** Ultrasonic Sensors and Methods, Fraunhofer IKTS, Dresden, Germany

**Martin H. Schulze** Eddy Current Methods, Fraunhofer IKTS, Dresden, Germany

**Gunnar Suchanek** Institute of Solid State Electronics, Technische Universität Dresden, Dresden, Germany

**Klaus Szielasko** Fraunhofer Institute for Nondestructive Testing IZFP, Saarbrücken, Germany

**Antonello Tamburrino** Department of Electrical and Information Engineering, Università degli Studi di Cassino e del Lazio Meridionale, Cassino, Italy

Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA

**Brian K. Tanner** Department of Physics, Durham University, Durham, UK

**João Manuel R. S. Tavares** Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

**Wolfgang Treimer** Department Mathematics-Physics-Chemistry, University of Applied Sciences Beuth Hochschule für Technik Berlin, Berlin, Germany

**Maciej Trusiak** Department of Mechatronics, Institute of Micromechanics and Photonics, Warsaw University of Technology, Warsaw, Poland

**Eleni Tsangouri** Department of Mechanics of Materials and Constructions (MeMC), Vrije Universiteit Brussel (VUB), Brussels, Belgium

**Kilian Tschöke** Fraunhofer Institute for Ceramic Technologies and Systems IKTS, Dresden, Germany

**Ralf Tschuncky** Fraunhofer Institute for Nondestructive Testing IZFP, Saarbrücken, Germany

**Bart Van Damme** Empa, Swiss Federal Laboratories for Materials Science and Technology, Dubendorf, Switzerland

**Vladimir P. Vavilov** Tomsk Polytechnic University, Tomsk, Russia

**Masoud Vejdannik** School of Electrical Engineering, Iran University of Science and Technology (IUST), Narmak, Tehran, Iran

**Julia Walther** Clinical Sensoring and Monitoring, Anesthesiology and Intensive Care Medicine, Faculty of Medicine Carl Gustav Carus, TU Dresden, Dresden, Germany

**Bianca Wehnacht** Fraunhofer Institute for Ceramic Technologies and Systems IKTS, Dresden, Germany

**Shejuan Xie** State Key Laboratory for Strength and Vibration of Mechanical Structures Shaanxi ERC for NDT and Structural Integrity Evaluation School of Aerospace, Xi'an Jiaotong University, Xian, China

**Lianxiang Yang** Optical Laboratory, Department of Mechanical Engineering, Oakland University, Rochester, MI, USA

**Qiwen Zhan** School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China

Department of Electro-Optics and Photonics, University of Dayton, Dayton, OH, USA

**Qixin Zhou** Department of Chemical and Biomolecular Engineering, Corrosion Engineering Program, The University of Akron, Akron, OH, USA

**Marek Ziolkowski** Technische Universitaet Ilmenau, Ilmenau, Germany

West Pomeranian University of Technology Szczecin, Szczecin, Poland

---

## **Part I**

### **Acoustic Techniques**



# Physical Basis for Ultrasonic Acoustics

1

Victor Giurgiutiu and Bin Lin

## Contents

Bulk Waves in an Infinite Elastic Medium .....	4
Plane Bulk Waves .....	4
Bulk Wave Types .....	6
Wave Potentials .....	8
Dilatational and Rotational Waves .....	9
Irrotational and Equivolume Waves .....	10
z-Invariant Waves .....	12
Bulk Waves at Bimaterial Interface .....	16
Solid-Solid Interface .....	17
General Solution .....	20
P-wave Excitation .....	23
SV-wave Excitation .....	25
Bulk Waves at the Interface Between Liquid and Solid Media .....	28
Bulk Waves in Liquids .....	28
Interface Between Liquid and Solid Media .....	31
Bulk Waves at Liquid-Solid Interface .....	32
Bulk Waves at Solid-Liquid Interface .....	35
References .....	41

## Abstract

This chapter presents a review of the physical basis of ultrasonic waves in elastic media and bimaterial interfaces. The chapter gives an overview of the description and presentation of the wave propagation problem, which can become quite

---

V. Giurgiutiu (✉)

University of South Carolina, Columbia, SC, USA  
e-mail: [VICTORG@sc.edu](mailto:VICTORG@sc.edu)

B. Lin

Department of Mechanical Engineering, University of South Carolina, Columbia, SC, USA  
e-mail: [LINBIN@cec.sc.edu](mailto:LINBIN@cec.sc.edu)

complicated in some cases. The chapter starts with discussion of the wave propagation equation in an infinite elastic medium. The general equations of 3-D wave propagation in unbounded solid media are developed from first principles. The eigenvalues and eigenvectors of the wave equation are identified. The two corresponding basic wave types, *pressure waves* and *shear waves*, are discussed. Dilatational, rotational, irrotational, and equivolume waves are identified and discussed. The case of  $z$ -invariant wave propagation is presented.

Bulk wave interaction at bimaterial interfaces is presented next. The solid-solid interface is studied first. The interface conditions are set up in terms of potentials and the coherence condition (aka Snell's law for ultrasonics) is derived. The general solution is deduced under the assumption that the incident P and SV waves are coherent with each other. When the P and SV waves are not coherent to each other, the general solution is made specific to separate P-wave and SV-wave situations. Critical angles for P and SV excitation are derived. The interface between liquid and solid media is treated next. Both liquid-solid (LS) and solid-liquid (SL) interfaces are separately considered. Several numerical examples are presented to illustrate the critical angle concept.

## Bulk Waves in an Infinite Elastic Medium

### Plane Bulk Waves

The equation describing waves propagating in an unbound elastic solid, aka *wave equation*, is

$$\nabla^2 \Phi(\vec{r}, t) = \frac{1}{c^2} \ddot{\Phi}(\vec{r}, t) \quad (\text{wave equation}) \quad (1)$$

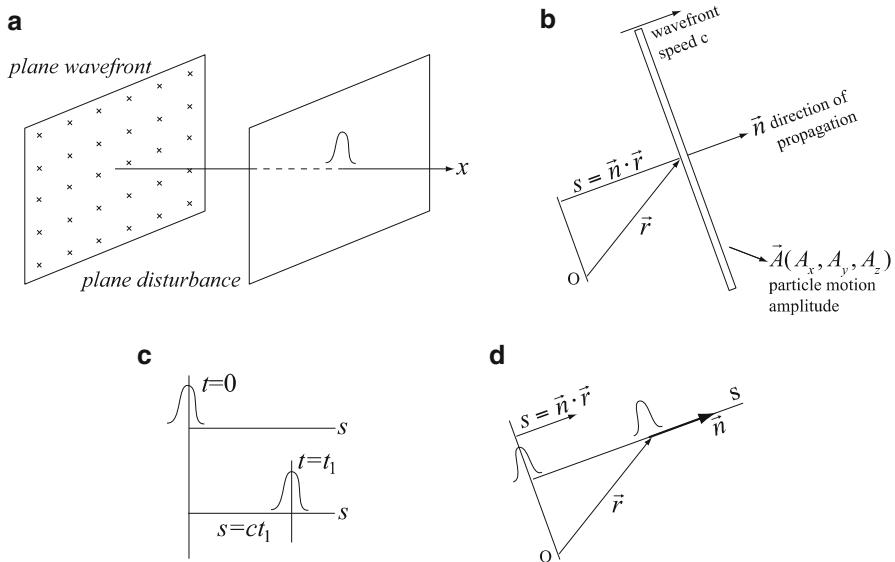
where  $\Phi(\vec{r}, t)$  is a generic disturbance propagating in the unbound elastic solid.

Plane waves are waves having a planar wave front. Plane waves may be generated by a planar source. All particles in the wave front are contained in a plane with normal unit vector  $\vec{n} = \{n_x, n_y, n_z\}$ . The planar wave front moves along  $\vec{n}$  with wave speed  $c$ . As the wave front propagates, it traces a path  $s$  (Fig. 1). The wave motion of the particles inside the planar wave front is self-similar and the Laplacian of  $\Phi$  along the path  $s$  in the direction  $\vec{n}$  is

$$\nabla^2 \Phi = \frac{\partial^2 \Phi}{\partial s^2} \quad (2)$$

Substitution of Eq. (2) into Eq. (1) yields the wave equation for plane waves, i.e.,

$$\frac{\partial^2 \Phi}{\partial s^2} = \frac{1}{c^2} \frac{\partial^2 \Phi}{\partial t^2} \quad (\text{plane wave equation}) \quad (3)$$



**Fig. 1** Plane waves: (a) planar disturbance generating a planar wave front; (b) the geometry of a planar wave front with normal  $\vec{n}$ ; (c) generic wave shape; (d) wave propagation path  $s$  measured along  $\vec{n}$

The d'Alembert solution of Eq. (3) with only retaining the term  $f(s - ct)$  represents a forward propagating wave, i.e.,

$$\Phi(s, t) = f(s - ct) \quad (\text{generic plane wave}) \quad (4)$$

The path  $s$  is the projection of the position vector  $\vec{r}$  onto the propagation direction  $\vec{n} = n_x \vec{i} + n_y \vec{j} + n_z \vec{k}$ , i.e.,

$$s = \vec{n} \cdot \vec{r} = n_x x + n_y y + n_z z \quad (5)$$

The particle motion  $\vec{u} = u_x \vec{i} + u_y \vec{j} + u_z \vec{k}$  has the general expression

$$\vec{u} = \vec{A} f(\vec{n} \cdot \vec{r} - ct) \quad (6)$$

where  $\vec{A} = A_x \vec{i} + A_y \vec{j} + A_z \vec{k}$  is the motion amplitude.

For harmonic plane waves, the forward propagation wave is

$$\Phi(\vec{r}, t) = A e^{i(\vec{\gamma} \cdot \vec{r} - \omega t)} = A e^{i(\xi x + \eta y + \zeta z - \omega t)} \quad (\text{harmonic plane wave}) \quad (7)$$

where  $\vec{\gamma} = \xi \vec{i} + \eta \vec{j} + \zeta \vec{k}$  is the wavevector and  $\gamma = |\vec{\gamma}|$  is the wavenumber. Note that

$$\xi = \gamma n_x, \quad \eta = \gamma n_y, \quad \zeta = \gamma n_z \quad (\text{directional wave numbers}) \quad (8)$$

The particle motion of a harmonic plane wave has the general expression

$$\vec{u}(\vec{r}, t) = \vec{A} e^{i(\vec{k} \cdot \vec{r} - \omega t)} = \vec{A} e^{i(\xi x + \eta y + \zeta z - \omega t)} \quad (\text{harmonic plane wave}) \quad (9)$$

## Bulk Wave Types

Recall the Navier-Lame governing equations in  $Ox_1x_2x_3$  Cartesian coordinates, i.e.,

$$\begin{aligned} (\lambda + \mu) \left( \frac{\partial^2 u_1}{\partial x_1^2} + \frac{\partial^2 u_2}{\partial x_1 \partial x_2} + \frac{\partial^2 u_3}{\partial x_1 \partial x_3} \right) + \mu \left( \frac{\partial^2 u_1}{\partial x_2^2} + \frac{\partial^2 u_2}{\partial x_2 \partial x_3} + \frac{\partial^2 u_3}{\partial x_2^2} \right) &= \rho \ddot{u}_1 \\ (\lambda + \mu) \left( \frac{\partial^2 u_1}{\partial x_2 \partial x_1} + \frac{\partial^2 u_2}{\partial x_2^2} + \frac{\partial^2 u_3}{\partial x_2 \partial x_3} \right) + \mu \left( \frac{\partial^2 u_2}{\partial x_1^2} + \frac{\partial^2 u_2}{\partial x_1 \partial x_3} + \frac{\partial^2 u_3}{\partial x_1^2} \right) &= \rho \ddot{u}_2 \\ (\lambda + \mu) \left( \frac{\partial^2 u_1}{\partial x_3 \partial x_1} + \frac{\partial^2 u_2}{\partial x_3 \partial x_2} + \frac{\partial^2 u_3}{\partial x_3^2} \right) + \mu \left( \frac{\partial^2 u_3}{\partial x_1^2} + \frac{\partial^2 u_3}{\partial x_1 \partial x_2} + \frac{\partial^2 u_3}{\partial x_2^2} \right) &= \rho \ddot{u}_3 \end{aligned} \quad (10)$$

where  $\lambda, \mu$  are the Lame elastic constants. Assume a generic plane wave propagating forward along an arbitrary direction  $\vec{n} = n_1 \vec{e}_1 + n_2 \vec{e}_2 + n_3 \vec{e}_3$ . Substitution of Eq. (6) into Eq. (10) yields, upon rearrangement,

$$\begin{bmatrix} (\lambda + \mu)n_1^2 + (\mu - \rho c^2) & (\lambda + \mu)n_2n_1 & (\lambda + \mu)n_3n_1 \\ (\lambda + \mu)n_1n_2 & (\lambda + \mu)n_2^2 + (\mu - \rho c^2) & (\lambda + \mu)n_3n_2 \\ (\lambda + \mu)n_1n_3 & (\lambda + \mu)n_2n_3 & (\lambda + \mu)n_3^2 + (\mu - \rho c^2) \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (11)$$

The homogeneous linear system of Eq. (11) accepts nontrivial solution only if its determinant vanishes. Imposition of this condition yields the characteristic equation

$$[(\lambda + 2\mu) - \rho c^2](\mu - \rho c^2)^2 = 0 \quad (12)$$

It is apparent that Eq. (12) accepts three solutions (eigenvalues), of which two are identical, i.e.,

$$c_I = \sqrt{\frac{\lambda + 2\mu}{\rho}}, \quad c_{II} = c_{III} = \sqrt{\frac{\mu}{\rho}} \quad (13)$$

Substitution of the eigenvalues of Eq. (13) into Eq. (11) yields the eigenvectors of the algebraic linear system (11). These eigenvectors are mutually orthogonal. The eigenvectors represent the wave polarization, i.e., the particle motion amplitude vector  $\vec{A} = A_x \vec{i} + A_y \vec{j} + A_z \vec{k}$ .

It is found that the eigenvector  $\vec{A}_I$  associated with the eigenvalue  $c_I$  has a particle motion which is parallel with the direction of propagation,  $\vec{n}$ . It is also found that the other two eigenvectors,  $\vec{A}_{II}, \vec{A}_{III}$ , which are associated with eigenvalues  $c_{II}, c_{III}$ , have particle motions which are transverse to the direction of propagation  $\vec{n}$  while also being mutually orthogonal. For illustration, if we assume that the  $x_1$  direction is chosen parallel to the direction of wave propagation, i.e.,  $\vec{n} = \vec{e}_1$ , then the three eigenvectors, i.e., polarization directions, are

$$\vec{A}_I = \vec{e}_1, \quad \vec{A}_{II} = \vec{e}_2, \quad \vec{A}_{III} = \vec{e}_3 \quad (\text{polarization directions}) \quad (14)$$

The waves with particle motion parallel to the direction of wave propagation are called *pressure waves* or simply *P-waves*. These waves are also known by other names such as *compressional, axial, dilatational, or longitudinal waves*. They travel with wave speed  $c_P$  given by

$$c_P = \sqrt{\frac{\lambda + 2\mu}{\rho}}, \quad c_P^2 = \frac{\lambda + 2\mu}{\rho} \quad (\text{pressure wave speed}) \quad (15)$$

The waves with particle motion perpendicular to the direction of wave propagation are called *shear waves* or simply *S-waves*. These waves are also known by other names such as *transverse waves* or *distortional waves*. They travel with wave speed  $c_S$  given by

$$c_S = \sqrt{\frac{\mu}{\rho}}, \quad c_S^2 = \frac{\mu}{\rho} \quad (\text{shear wave speed}) \quad (16)$$

Note that some authors use  $c_1$  and  $c_2$  instead of  $c_P$  and  $c_S$ , whereas other authors use  $c_L$  and  $c_T$ , where L and T stand for longitudinal and transverse, respectively.

The pressure and shear wave speeds  $c_P$  and  $c_S$  can be also expressed in engineering notations,  $E, G, v$ , i.e.,

$$c_P = \sqrt{\frac{1 - \nu}{(1 + \nu)(1 - 2\nu)} \frac{E}{\rho}}, \quad c_S = \sqrt{\frac{G}{\rho}} = \sqrt{\frac{1}{2(1 + \nu)} \frac{E}{\rho}} \quad (17)$$

In view of Eqs. (13), (14), (15), and (16), it is apparent that, in an unbound elastic solid, three possible bulk waves may exist, one pressure wave and two shear waves, i.e.,

$$\vec{u}_P = A_P \vec{e}_1 f(x_1 - c_P t) \quad (\text{pressure wave, P}) \quad (18)$$

$$\vec{u}_{SH} = A_{SH} \vec{e}_2 f(x_1 - c_S t) \quad (\text{shear-horizontal wave, SH}) \quad (19)$$

$$\vec{u}_{SV} = A_{SV} \vec{e}_3 f(x_1 - c_S t) \quad (\text{shear-vertical wave, SV}) \quad (20)$$

where  $A_P, A_{SH}, A_{SV}$  are the wave amplitudes. The *total wave solution*  $\vec{u}$  results from the superposition of the individual solutions, i.e.,

$$\vec{u} = \vec{u}_P + \vec{u}_{SH} + \vec{u}_{SV} \quad (\text{total wave}) \quad (21)$$

For a backward propagating wave, the same approach can be applied.

In the case of *harmonic waves* as described by Eq. (9), one gets

$$\vec{u}_P = A_P \vec{e}_1 e^{i(\gamma_P x_1 - \omega t)} \quad (\text{pressure wave}) \quad (22)$$

$$\vec{u}_{SH} = A_{SH} \vec{e}_2 e^{i(\gamma_S x_1 - \omega t)} \quad (\text{shear-horizontal wave}) \quad (23)$$

$$\vec{u}_{SV} = A_{SV} \vec{e}_3 e^{i(\gamma_S x_1 - \omega t)} \quad (\text{shear-vertical wave}) \quad (24)$$

The wavenumbers  $\gamma_P, \gamma_S$ , in Eqs. (22), (23), and (24) are given by

$$\gamma_P = \frac{\omega}{c_P}, \quad \gamma_S = \frac{\omega}{c_S} \quad (25)$$

Note that Eqs. (18), (19), (20), (21), (22), (23), and (24) assume that the  $x_1$  direction was chosen parallel to the direction of wave propagation, i.e.,  $\vec{n} = \vec{e}_1$ . In the general case when  $\vec{n} = n_1 \vec{e}_1 + n_2 \vec{e}_2 + n_3 \vec{e}_3$ , appropriate rotation of axes must be applied; however, the intrinsic physical properties of the P, SH, and SV waves will be the same.

## Wave Potentials

Using the Helmholtz decomposition concept, assume that the displacement  $\vec{u}$  can be expressed in terms of two potential functions, a *scalar potential*  $\Phi$  and a *vector potential*  $\vec{H} = H_x \vec{i} + H_y \vec{j} + H_z \vec{k}$ , i.e.,

$$\begin{aligned} \vec{u} &= \text{grad } \Phi + \text{curl } \vec{H} = \vec{\nabla} \Phi + \vec{\nabla} \times \\ &\quad \vec{H} \quad (\text{displacement in terms of potentials}) \end{aligned} \quad (26)$$

Equation (26) is complemented by the *uniqueness condition*

$$\vec{\nabla} \cdot \vec{H} = 0 \quad (\text{uniqueness condition}) \quad (27)$$

Recall the Navier-Lame equations in vector form as

$$(\lambda + \mu) \vec{\nabla} \left( \vec{\nabla} \cdot \vec{u} \right) + \mu \vec{\nabla}^2 \vec{u} = \rho \ddot{\vec{u}} \quad (28)$$

The *wave equations* for the scalar potential  $\Phi$  and the vector potential  $\vec{H}$ , i.e.,

$$c_P^2 \vec{\nabla}^2 \Phi = \ddot{\Phi} \quad (\text{wave equation for scalar potential } \Phi) \quad (29)$$

$$c_S^2 \vec{\nabla}^2 \vec{H} = \ddot{\vec{H}} \quad (\text{wave equation for vector potential } \vec{H}) \quad (30)$$

Equation (29) indicates that the scalar potential,  $\Phi$ , propagates with the pressure wave speed,  $c_P$ , whereas Eq. (30) indicates that the vector potential,  $\vec{H}$ , propagates with the shear wave speed,  $c_S$ .

## Dilatational and Rotational Waves

Dilatational waves involve the dilation  $\Delta = \vec{\nabla} \cdot \vec{u}$ , whereas rotational waves involve the rotation  $\vec{\omega} = \frac{1}{2} \vec{\nabla} \times \vec{u}$ . Dilation  $\Delta$  is a scalar representing increase in volume. Rotation  $\vec{\omega}$  is a vector having as components the rotations about each of the coordinate axes.

To calculate the dilatational waves, recall the Navier-Lame equations in vector form as

$$(\lambda + 2\mu) \vec{\nabla} \Delta - 2\mu \vec{\nabla} \times \vec{\omega} = \rho \ddot{\vec{u}} \quad (31)$$

Dot premultiply Eq. (31) by  $\vec{\nabla} \cdot$  and get

$$(\lambda + 2\mu) \vec{\nabla} \cdot \vec{\nabla} \Delta - 2\mu \vec{\nabla} \cdot (\vec{\nabla} \times \vec{\omega}) = \rho \vec{\nabla} \cdot \ddot{\vec{u}} \quad (32)$$

Upon simplification, Eq. (32) yields the *dilatation wave equation*, i.e.,

$$(\lambda + 2\mu) \nabla^2 \Delta = \rho \ddot{\Delta} \quad (33)$$

where  $\nabla^2 = \vec{\nabla} \cdot \vec{\nabla}$ . The simplification of Eq. (32) into Eq. (33) was made possible by the general vector property  $\vec{a} \cdot (\vec{a} \times \vec{b}) = 0$  and by the inter-commutability of space and time derivatives. Division of Eq. (33) by  $\rho$  and utilization of Eq. (15) yields

$$c_P^2 \nabla^2 \Delta = \ddot{\Delta} \quad (34)$$

Equation (34) indicates that dilatational waves propagate with the pressure wave speed,  $c_P$ .

To calculate the rotational waves, recall the Navier-Lame equations in vector form as

$$(\lambda + \mu) \vec{\nabla} \Delta + \mu \nabla^2 \vec{u} = \rho \ddot{\vec{u}} \quad (35)$$

Cross premultiply Eq. (35) by  $\vec{\nabla} \times$  and get

$$(\lambda + 2\mu) \vec{\nabla} \times \vec{\nabla} \Delta + \mu \nabla^2 \vec{\nabla} \times \vec{u} = \rho \vec{\nabla} \times \ddot{\vec{u}} \quad (36)$$

Upon simplification, Eq. (36) yields the *rotational wave equation*, i.e.,

$$\mu \nabla^2 \vec{\omega} = \rho \ddot{\vec{\omega}} \quad (37)$$

The simplification of Eq. (36) into Eq. (37) was made possible by the vector property  $\vec{a} \times \vec{a} = \vec{0}$  and by the inter-commutability of space and time derivatives. Division of Eq. (33) by  $\rho$  and utilization of Eq. (16) yields

$$c_s^2 \nabla^2 \vec{\omega} = \ddot{\vec{\omega}} \quad (38)$$

Equation (38) indicates that rotational waves propagate with the shear wave speed,  $c_s$ .

Equations (33) and (37) indicate that both dilatation and rotation obey the wave equation and that they propagate as waves into the elastic body. Depending on the initial and boundary conditions, the dilatational and rotational waves can either exist alone or may coexist.

## Irrational and Equivolume Waves

*Irrational waves* have zero rotation, i.e.,  $\vec{\omega} = 0$ , whereas *equivolume waves* have zero dilatation, i.e.,  $\Delta = \vec{\nabla} \cdot \vec{u} = 0$ .

*Pressure waves are irrational.* The proof is as follows: Assume a pressure wave in the form of Eq. (18), i.e.,

$$\vec{u} = A_P \vec{e}_1 f(x_1 - c_P t) \quad (\text{pressure wave}) \quad (39)$$

Calculate the rotation and show that it is zero, i.e.,

$$\begin{aligned} \vec{\omega} &= \frac{1}{2} \vec{\nabla} \times \vec{u} = \frac{1}{2} \vec{\nabla} \times A_P \vec{e}_1 f(x_1 - c_P t) \\ &= \frac{1}{2} A_P \left( \frac{\partial}{\partial x_3} \vec{e}_2 - \frac{\partial}{\partial x_2} \vec{e}_3 \right) f(x_1 - c_P t) = \vec{0} \end{aligned} \quad (40)$$

This proves that *pressure waves are irrational waves*, i.e., it has zero rotation. The dilatation is calculated as

$$\Delta = \vec{\nabla} \cdot \vec{u} = \vec{\nabla} \cdot A_P \vec{e}_1 f(x_1 - c_P t) = A_P \frac{\partial}{\partial x_1} f(x_1 - c_P t) = A_P f'(x_1 - c_P t) \neq 0 \quad (41)$$

This proves that the *pressure waves are dilatational waves*. However, the shear strains are nonzero, e.g.,

$$\varepsilon_{12} = \frac{1}{2} \left( \frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \right) = \frac{1}{2} A_P f'(x_1 - c_P t) \neq 0 \quad (42)$$

This means that the pressure waves are *not* nondistortional.

In terms of wave potentials formulation, we note that waves based only on the scalar potential,  $\Phi$ , are irrotational waves. To prove this, assume

$$\vec{u} = \vec{\nabla} \Phi \quad (43)$$

and calculate the rotation

$$\vec{\omega} = \frac{1}{2} \vec{\nabla} \times \vec{u} = \frac{1}{2} (\vec{\nabla} \times \vec{\nabla}) \Phi = \vec{0} \quad (44)$$

Equation (44) is zero in virtue of the vector property  $\vec{a} \times \vec{a} = \vec{0}$ . On the other hand, the dilatation is given by

$$\Delta = \vec{\nabla} \cdot \vec{u} = \vec{\nabla} \cdot \vec{\nabla} \Phi = \nabla^2 \Phi \neq 0 \quad (45)$$

This implies that the scalar potential,  $\Phi$ , can be also seen as the *dilatation potential*.

*Shear waves are equivolume.* The proof is as follows: Assume a shear wave in the form of, say, Eq. (19), i.e.,

$$\vec{u} = A_{SH} \vec{e}_2 f(x_1 - c_{st}) \quad (\text{shear wave}) \quad (46)$$

Calculate the dilatation and show that it is zero, i.e.,

$$\Delta = \vec{\nabla} \cdot \vec{u} = \vec{\nabla} \cdot A_{SH} \vec{e}_2 f(x_1 - c_{st}) = A_{SH} \frac{\partial}{\partial x_2} f(x_1 - c_{st}) = 0 \quad (47)$$

Recall that the dilatation is a direct measure of volumetric change; hence, zero dilatation implies a zero change of volume. This proves that the shear waves are *equivolume waves*. On physical terms, it is apparent that during shearing, the only deformation that the medium undergoes is distortion. Distortion takes place without any change of volume. This also implies that the equivolume waves are also *distortional waves*.

In terms of wave potentials formulation, we notice that waves based only on the vector potential  $\vec{H}$  are equivolume waves. To prove this, assume a wave depending only on  $\vec{H}$ , i.e.,

$$\vec{u} = \vec{\nabla} \times \vec{H} \quad (48)$$

and calculate the dilatation

$$\Delta = \vec{\nabla} \cdot \vec{u} = \vec{\nabla} \cdot (\vec{\nabla} \times \vec{H}) = 0 \quad (49)$$

Equation (49) is zero in virtue of the vector property  $\vec{a} \cdot (\vec{a} \times \vec{b}) = 0$ .

The role of  $\vec{H}$  potential is both distortional and rotational. The distortional role stems from the fact that a wave base on  $\vec{H}$  is equivolume as indicated by Eq. (49). Since distortion happens without volume change, it follows that the  $\vec{H}$  potential is responsible for the distortional component of the wave motion. The rotational role of  $\vec{H}$  stems from the following calculation of the rotation  $\vec{\omega}$ , i.e.,

$$\vec{\omega} = \frac{1}{2} \vec{\nabla} \times \vec{u} = \frac{1}{2} \vec{\nabla} \times (\vec{\nabla} \Phi + \vec{\nabla} \times \vec{H}) = \frac{1}{2} \cancel{\vec{\nabla} \times \vec{\nabla} \Phi} + \frac{1}{2} \vec{\nabla} \times \vec{\nabla} \times \vec{H} \neq 0 \quad (50)$$

It is apparent from Eq. (50) that  $\vec{H}$  generates a nonzero rotation. Thus, the vector potential,  $\vec{H}$ , can be seen as a *distortional potential* as well as a *rotation potential*.

An interesting note to be made is that while distortional (shear) waves are nondilatational, the dilatational (pressure) waves are not necessarily nondistortional. In fact, dilatational waves contain both bulk expansion and distortion. A simple example is provided by the case of spherical waves emanating from a point-size blast. By virtue of spherical symmetry, the motion is clearly irrotational. However, examination of an infinitesimal element indicates that its deformation is not solely volumetric. In fact, in order to achieve radial expansion, both bulk expansion and distortion (shear strains) are necessary. This observation is substantiated numerically by the relation between the constant  $\lambda+2\mu$ , on one hand, and the bulk modulus,  $B$ , and shear modulus,  $\mu$ , on the other hand, i.e.,

$$\lambda + 2\mu = B + \frac{4}{3}\mu \quad (51)$$

It is apparent that the constant  $\lambda+2\mu$ , which defines pressure wave speed  $c_P$  via Eq. (15), depends on both the bulk modulus,  $B$ , and the shear modulus,  $\mu$ . This means that irrotational waves involve both bulk expansion and distortion.

## **z-Invariant Waves**

Of particular interest are 3-D plane waves that are invariant in one direction along the wave front. This situation is encountered, for example, in straight-crested waves, with the wave crest parallel to the  $z$ -direction. Usually, the invariant direction is taken as the  $z$ -axis, and for this reason these waves are called *z-invariant*. Under the  $z$ -invariant conditions, certain simplifications apply that are of particular usefulness in ultrasonic NDE practice.

For convenience, revert from  $x_1, x_2, x_3$  notations to  $x, y, z$  notations. Assume the wave front is parallel to the  $z$ -axis, and that the wave disturbance is invariant along the  $z$ -axis (Fig. 2). This means that all the functions involved in the analysis will not depend on  $z$ , and their derivatives with respect to  $z$  will be zero. In addition, the wave

front direction,  $\vec{n}$ , will be perpendicular to the  $z$ -axis, i.e.,  $\vec{n} \perp \vec{k}$ , where  $\vec{k}$  is the unit vector of the  $z$ -axis. Recall Eq. (26), i.e.,

$$\vec{u} = \vec{\nabla}\Phi + \vec{\nabla} \times \vec{H} \quad (52)$$

The potentials  $\Phi$ ,  $H_x$ ,  $H_y$ ,  $H_z$  satisfy the wave equations and the uniqueness condition, i.e.,

$$c_p^2 \nabla^2 \Phi = \ddot{\Phi}, \quad \begin{cases} c_s^2 \nabla^2 H_x = \ddot{H}_x \\ c_s^2 \nabla^2 H_y = \ddot{H}_y, \quad \frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z} = 0 \\ c_s^2 \nabla^2 H_z = \ddot{H}_z \end{cases} \quad (53)$$

The  $z$ -invariant condition yields

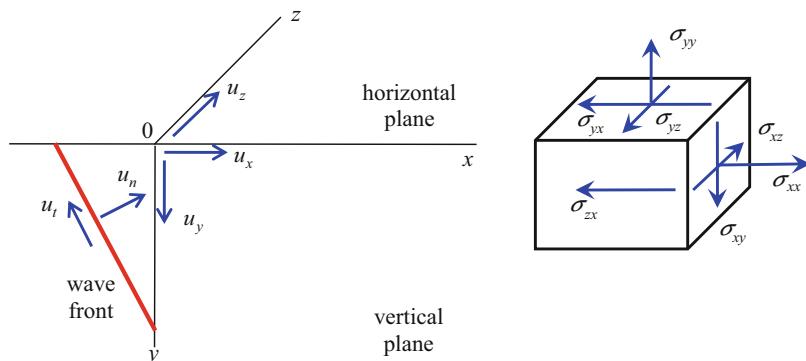
$$\frac{\partial}{\partial z} \equiv 0 \quad \text{and} \quad \vec{\nabla} = \vec{i} \frac{\partial}{\partial x} + \vec{j} \frac{\partial}{\partial y} \quad (54)$$

The use of Eq. (54) into Eq. (52) yields, upon expansion,

$$\vec{u} = \left( \frac{\partial \Phi}{\partial x} + \frac{\partial H_z}{\partial y} \right) \vec{i} + \left( \frac{\partial \Phi}{\partial y} - \frac{\partial H_z}{\partial x} \right) \vec{j} + \left( \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \right) \vec{k} \quad (55)$$

Equation (55) indicates that, though the motion is  $z$ -invariant, the displacement has components in all three directions ( $x$ ,  $y$ ,  $z$ ), i.e.,

$$u_x = \frac{\partial \Phi}{\partial x} + \frac{\partial H_z}{\partial y}, \quad u_y = \frac{\partial \Phi}{\partial y} - \frac{\partial H_z}{\partial x}, \quad u_z = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \quad (56)$$



**Fig. 2** The general setup for the study of  $z$ -invariant plane waves

Recall the stress-displacement relations, i.e.,

$$\begin{aligned}\sigma_{xx} &= (\lambda + 2\mu) \frac{\partial u_x}{\partial x} + \lambda \frac{\partial u_y}{\partial y} + \lambda \frac{\partial u_z}{\partial z} & \sigma_{xy} &= \mu \left( \frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} \right) \\ \sigma_{yy} &= \lambda \frac{\partial u_x}{\partial x} + (\lambda + 2\mu) \frac{\partial u_y}{\partial y} + \lambda \frac{\partial u_z}{\partial z} & \text{and} & \sigma_{yz} = \mu \left( \frac{\partial u_y}{\partial z} + \frac{\partial u_z}{\partial y} \right) \\ \sigma_{zz} &= \lambda \frac{\partial u_x}{\partial x} + \lambda \frac{\partial u_y}{\partial y} + (\lambda + 2\mu) \frac{\partial u_z}{\partial z} & \sigma_{zx} &= \mu \left( \frac{\partial u_z}{\partial x} + \frac{\partial u_x}{\partial z} \right)\end{aligned}\quad (57)$$

The use of Eq. (54) into Eq. (57) yields

$$\begin{aligned}\sigma_{xx} &= (\lambda + 2\mu) \frac{\partial u_x}{\partial x} + \lambda \frac{\partial u_y}{\partial y} & \sigma_{xy} &= \mu \left( \frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} \right) \\ \sigma_{yy} &= \lambda \frac{\partial u_x}{\partial x} + (\lambda + 2\mu) \frac{\partial u_y}{\partial y} & \text{and} & \sigma_{yz} = \mu \frac{\partial u_z}{\partial y} \\ \sigma_{zz} &= \lambda \frac{\partial u_x}{\partial x} + \lambda \frac{\partial u_y}{\partial y} & \sigma_{zx} &= \mu \frac{\partial u_z}{\partial x}\end{aligned}\quad (58)$$

Examinations of Eq. (56) indicate that it is possible to partition the solution into two parts: (a) a solution for  $u_z$  which depends only on the two potentials,  $H_x$  and  $H_y$ , and (b) a separate solution for  $u_x$  and  $u_y$  which depend on the other two potentials,  $\Phi$  and  $H_z$ .

The first solution, which accepts only the  $u_z$  displacement, will be a shear motion polarized in the horizontal plane  $Oxz$ , i.e., a shear-horizontal (SH) wave. This SH motion is described in terms of the two potentials,  $H_x$  and  $H_y$ . Examination of Eq. (58) reveals that the only nonzero stress components of this solution are the shear stresses  $\sigma_{yz}, \sigma_{xz}$ , because these are the only stresses that depend on the  $z$ -displacement  $u_z$ .

The second solution, which accepts  $u_x$  and  $u_y$  displacements, will be the combination of a pressure (P) wave represented by the potential  $\Phi$  and a shear vertical (SV) wave represented by the potential  $H_z$ . This second solution is denoted as P+SV. Note that the particle motion of this second solution is constrained into the vertical plane  $Oxy$ ; hence, the associated shear wave is a shear vertical wave, SV. The two solutions are treated separately. Examination of Eq. (58) reveals that the nonzero stress components of this solution are  $\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{xy}, \sigma_{yz}, \sigma_{xz}$ , because these are the stresses that depend on the displacements  $u_x, u_y$ . We will analyze now these two solutions separately as follows.

**SH wave solution:** the motion is contained in the horizontal plane, and the relevant potentials are  $H_x$  and  $H_y$ , i.e.,

$$u_x = u_y = 0, \quad u_z \neq 0, \quad \frac{\partial}{\partial z} = 0 \quad H_x \quad \text{and} \quad H_y \quad \text{only} \quad (59)$$

The stresses Eq. (58) simplifies to

$$\left. \begin{array}{l} \sigma_{xx} = 0 \\ \sigma_{yy} = 0 \\ \sigma_{zz} = 0 \end{array} \right| \quad \left. \begin{array}{l} \sigma_{xy} = 0 \\ \sigma_{yz} = \mu \frac{\partial u_z}{\partial y} \\ \sigma_{zx} = \mu \frac{\partial u_z}{\partial x} \end{array} \right\} \quad (60)$$

The nonzero  $u_z$  displacement and its derivatives are given by

$$u_z = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}, \quad \frac{\partial u_z}{\partial x} = \frac{\partial^2 H_y}{\partial x^2} - \frac{\partial^2 H_x}{\partial x \partial y}, \quad \frac{\partial u_z}{\partial y} = \frac{\partial^2 H_y}{\partial x \partial y} - \frac{\partial^2 H_x}{\partial y^2} \quad (61)$$

Substitution of Eqs. (56) and (59) into the stress-displacement relations (60) yields

$$\sigma_{yz} = \mu \frac{\partial u_z}{\partial y} = \mu \left( -\frac{\partial^2 H_x}{\partial y^2} + \frac{\partial^2 H_y}{\partial x \partial y} \right) \quad (\text{SH waves}) \quad (62)$$

$$\sigma_{zx} = \mu \frac{\partial u_z}{\partial x} = \mu \left( -\frac{\partial^2 H_x}{\partial x \partial y} + \frac{\partial^2 H_y}{\partial x^2} \right) \quad (\text{SH waves}) \quad (63)$$

Note that all the other stresses are null, i.e.,  $\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = \sigma_{xy} = 0$

**P + SV wave solution:** motion is contained in the vertical plane, and the relevant potentials are  $\Phi$  and  $H_z$ , i.e.,

$$u_x \neq 0, \quad u_y \neq 0, \quad u_z = 0, \quad \frac{\partial}{\partial z} = 0, \quad \Phi \text{ and } H_z \text{ only} \quad (64)$$

The nonzero  $u_x$ ,  $u_y$  displacements and their derivatives are

$$\begin{aligned} u_x &= \frac{\partial \Phi}{\partial x} + \frac{\partial H_z}{\partial y}, & \frac{\partial u_x}{\partial x} &= \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 H_z}{\partial x \partial y}, & \frac{\partial u_x}{\partial y} &= \frac{\partial^2 \Phi}{\partial x \partial y} + \frac{\partial^2 H_z}{\partial y^2} \\ u_y &= \frac{\partial \Phi}{\partial y} - \frac{\partial H_z}{\partial x}, & \frac{\partial u_y}{\partial x} &= \frac{\partial^2 \Phi}{\partial x \partial y} - \frac{\partial^2 H_z}{\partial x^2}, & \frac{\partial u_y}{\partial y} &= \frac{\partial^2 \Phi}{\partial y^2} - \frac{\partial^2 H_z}{\partial x \partial y} \end{aligned} \quad (65)$$

Substitution of Eqs. (64) and (65) into Eq. (58) yields

$$\begin{aligned} \sigma_{xx} &= (\lambda + 2\mu) \left( \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 H_z}{\partial x \partial y} \right) + \lambda \left( \frac{\partial^2 \Phi}{\partial y^2} - \frac{\partial^2 H_z}{\partial x \partial y} \right) = (\lambda + 2\mu) \nabla^2 \Phi - 2\mu \frac{\partial^2 \Phi}{\partial y^2} + 2\mu \frac{\partial^2 H_z}{\partial x \partial y} \\ \sigma_{yy} &= \lambda \left( \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 H_z}{\partial x \partial y} \right) + (\lambda + 2\mu) \left( \frac{\partial^2 \Phi}{\partial y^2} - \frac{\partial^2 H_z}{\partial x \partial y} \right) = (\lambda + 2\mu) \nabla^2 \Phi - 2\mu \frac{\partial^2 \Phi}{\partial x^2} - 2\mu \frac{\partial^2 H_z}{\partial x \partial y} \\ \sigma_{zz} &= \lambda \left( \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 H_z}{\partial x \partial y} \right) + \lambda \left( \frac{\partial^2 \Phi}{\partial y^2} - \frac{\partial^2 H_z}{\partial x \partial y} \right) = \lambda \left( \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} \right) \\ \sigma_{xy} &= \mu \left( 2 \frac{\partial^2 \Phi}{\partial x \partial y} - \frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} \right) \quad (\text{P + SV waves}) \end{aligned} \quad (66)$$

where

$$\nabla^2 = \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} \quad (\text{P + SV waves}) \quad (67)$$

Note that, as expected,  $\sigma_{yz} = \sigma_{zx} = 0$ .

## Bulk Waves at Bimaterial Interface

Consider a bonded bi-material interface between two solid materials of properties  $\rho_1, \lambda_1, \mu_1, c_{P1}, c_{S1}$  and  $\rho_2, \lambda_2, \mu_2, c_{P2}, c_{S2}$ . Incident P and SV waves impinge obliquely on the interface at angles  $\theta_p^i, \theta_s^i$ , respectively (Fig. 3).

The potentials for the incident, reflected, and transmitted waves are:

$$\Phi^i = \hat{\Phi}^i e^{i(\xi_p^i x - \eta_p^i y - \omega t)}, \quad H_z^i = \hat{H}_z^i e^{i(\xi_s^i x - \eta_s^i y - \omega t)} \quad (\text{incident wave}) \quad (68)$$

$$\Phi^r = \hat{\Phi}^r e^{i(\xi_p^r x + \eta_p^r y - \omega t)}, \quad H_z^r = \hat{H}_z^r e^{i(\xi_s^r x + \eta_s^r y - \omega t)} \quad (\text{reflected wave}) \quad (69)$$

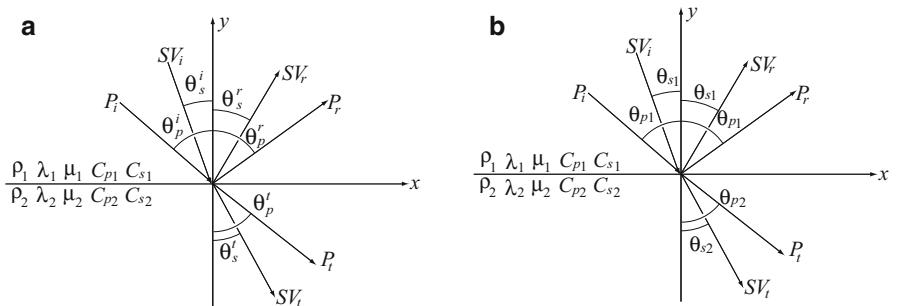
$$\Phi^t = \hat{\Phi}^t e^{i(\xi_p^t x - \eta_p^t y - \omega t)}, \quad H_z^t = \hat{H}_z^t e^{i(\xi_s^t x - \eta_s^t y - \omega t)} \quad (\text{transmitted wave}) \quad (70)$$

where

$$\xi_p^i = \gamma_{P1} \sin \theta_p^i \quad \eta_p^i = \gamma_{P1} \cos \theta_p^i \quad \gamma_{P1} = \omega / c_{P1}, \quad \text{etc.} \quad (71)$$

Substitution of Eqs. (68), (69), and (70) into Eq. (56) yields displacements for incident, reflected, and transmitted waves, i.e.,

$$u_x^i = i\xi_p^i \Phi^i - i\eta_p^i H_z^i, \quad u_x^r = i\xi_p^r \Phi^r + i\eta_p^r H_z^r, \quad u_x^t = i\xi_p^t \Phi^t - i\eta_p^t H_z^t \quad (72)$$



**Fig. 3** P and SV waves impinging obliquely on a bi-material interface: (a) initial setup; (b) simplified setup resulting from the application of the coherence condition

$$u_y^i = -i\eta_P^i \Phi^i - i\xi_S^i H_z^i, \quad u_y^r = i\eta_P^r \Phi^r - i\xi_S^r H_z^r, \quad u_y^t = -i\eta_P^t \Phi^t - i\xi_S^t H_z^t \quad (73)$$

Substitution of Eqs. (68), (69), and (70) into Eq. (66) yields stresses for incident, reflected, and transmitted waves, i.e.,

$$\sigma_{yy}^i = \left[ 2\mu_1 (\xi_P^i)^2 - (\lambda_1 + 2\mu_1)\gamma_{P1}^2 \right] \Phi^i - 2\mu_1 \xi_S^i \eta_S^i H_z^i \quad (74)$$

$$\sigma_{yy}^r = \left[ 2\mu_1 (\xi_P^r)^2 - (\lambda_1 + 2\mu_1)\gamma_{P1}^2 \right] \Phi^r + 2\mu_1 \xi_S^r \eta_S^r H_z^r \quad (75)$$

$$\sigma_{yy}^t = \left[ 2\mu_2 (\xi_P^t)^2 - (\lambda_2 + 2\mu_2)\gamma_{P2}^2 \right] \Phi^t - 2\mu_2 \xi_S^t \eta_S^t H_z^t \quad (76)$$

$$\sigma_{xy}^i = 2\mu_1 \xi_P^i \eta_P^i \Phi^i + \mu_1 \left[ (\xi_S^i)^2 - (\eta_S^i)^2 \right] H_z^i \quad (77)$$

$$\sigma_{xy}^r = -2\mu_1 \xi_P^r \eta_P^r \Phi^r + \mu_1 \left[ (\xi_S^r)^2 - (\eta_S^r)^2 \right] H_z^r \quad (78)$$

$$\sigma_{xy}^t = 2\mu_2 \xi_P^t \eta_P^t \Phi^t + \mu_2 \left[ (\xi_S^t)^2 - (\eta_S^t)^2 \right] H_z^t \quad (79)$$

## Solid-Solid Interface

Analysis of wave interaction at the interface is performed by setting  $y = 0$  in the potential, displacement, and stress expression and then imposing the displacement continuity and stress balance conditions.

### Potentials, Displacements, and Stresses at the Interface

At the interface  $y = 0$ , the potentials take the form

$$\Phi^i|_{y=0} = \hat{\Phi}^i e^{i(\xi_P^i x - \omega t)}, \quad H_z^i|_{y=0} = \hat{H}_z^i e^{i(\xi_S^i x - \omega t)} \quad (80)$$

$$\Phi^r|_{y=0} = \hat{\Phi}^r e^{i(\xi_P^r x - \omega t)}, \quad H_z^r|_{y=0} = \hat{H}_z^r e^{i(\xi_S^r x - \omega t)} \quad (81)$$

$$\Phi^t|_{y=0} = \hat{\Phi}^t e^{i(\xi_P^t x - \omega t)}, \quad H_z^t|_{y=0} = \hat{H}_z^t e^{i(\xi_S^t x - \omega t)} \quad (82)$$

Substitution of Eqs. (80), (81), and (82) into Eqs. (72), and (73) yields the displacements at the interface as

$$u_x^i|_{y=0} = i\xi_P^i \hat{\Phi}^i e^{i(\xi_P^i x - \omega t)} - i\eta_S^i \hat{H}_z^i e^{i(\xi_S^i x - \omega t)} \quad (83)$$

$$u_x^r|_{y=0} = i\xi_P^r \hat{\Phi}^r e^{i(\xi_P^r x - \omega t)} + i\eta_S^r \hat{H}_z^r e^{i(\xi_S^r x - \omega t)} \quad (84)$$

$$u_x^t|_{y=0} = i\xi_P^t \hat{\Phi}^t e^{i(\xi_P^t x - \omega t)} - i\eta_P^t \hat{H}_z^t e^{i(\xi_S^t x - \omega t)} \quad (85)$$

$$u_y^i|_{y=0} = -i\eta_P^i \hat{\Phi}^i e^{i(\xi_P^i x - \omega t)} - i\xi_S^i \hat{H}_z^i e^{i(\xi_S^i x - \omega t)} \quad (86)$$

$$u_y^r|_{y=0} = i\eta_P^r \hat{\Phi}^r e^{i(\xi_P^r x - \omega t)} - i\xi_S^r \hat{H}_z^r e^{i(\xi_S^r x - \omega t)} \quad (87)$$

$$u_y^t|_{y=0} = -i\eta_P^t \hat{\Phi}^t e^{i(\xi_P^t x - \omega t)} - i\xi_S^t \hat{H}_z^t e^{i(\xi_S^t x - \omega t)} \quad (88)$$

Substitution of Eqs. (80), (81), and (82) into Eqs. (74), (75), (76), (77), (78), and (79) yields the stresses at the interface as

$$\sigma_{yy}^i|_{y=0} = \left[ 2\mu_1 (\xi_P^i)^2 - (\lambda_1 + 2\mu_1) \gamma_{P1}^2 \right] \hat{\Phi}^i e^{i(\xi_P^i x - \omega t)} - 2\mu_1 \xi_S^i \eta_S^i \hat{H}_z^i e^{i(\xi_S^i x - \omega t)} \quad (89)$$

$$\sigma_{yy}^r|_{y=0} = \left[ 2\mu_1 (\xi_P^r)^2 - (\lambda_1 + 2\mu_1) \gamma_{P1}^2 \right] \hat{\Phi}^r e^{i(\xi_P^r x - \omega t)} + 2\mu_1 \xi_S^r \eta_S^r \hat{H}_z^r e^{i(\xi_S^r x - \omega t)} \quad (90)$$

$$\sigma_{yy}^t|_{y=0} = \left[ 2\mu_2 (\xi_P^t)^2 - (\lambda_2 + 2\mu_2) \gamma_{P2}^2 \right] \hat{\Phi}^t e^{i(\xi_P^t x - \omega t)} - 2\mu_2 \xi_S^t \eta_S^t \hat{H}_z^t e^{i(\xi_S^t x - \omega t)} \quad (91)$$

$$\sigma_{xy}^i|_{y=0} = 2\mu_1 \xi_P^i \eta_P^i \hat{\Phi}^i e^{i(\xi_P^i x - \omega t)} + \mu_1 \left[ (\xi_S^i)^2 - (\eta_S^i)^2 \right] \hat{H}_z^i e^{i(\xi_S^i x - \omega t)} \quad (92)$$

$$\sigma_{xy}^r|_{y=0} = -2\mu_1 \xi_P^r \eta_P^r \hat{\Phi}^r e^{i(\xi_P^r x - \omega t)} + \mu_1 \left[ (\xi_S^r)^2 - (\eta_S^r)^2 \right] \hat{H}_z^r e^{i(\xi_S^r x - \omega t)} \quad (93)$$

$$\sigma_{xy}^t|_{y=0} = 2\mu_2 \xi_P^t \eta_P^t \hat{\Phi}^t e^{i(\xi_P^t x - \omega t)} + \mu_2 \left[ (\xi_S^t)^2 - (\eta_S^t)^2 \right] \hat{H}_z^t e^{i(\xi_S^t x - \omega t)} \quad (94)$$

## Interface Conditions

Displacement compatibility at the interface is given by

$$u_{x1}|_{y=0} = u_{x2}|_{y=0}, \quad u_{y1}|_{y=0} = u_{y2}|_{y=0} \quad (95)$$

Traction balance at the interface is given by

$$\sigma_{yy1}|_{y=0} = \sigma_{yy2}|_{y=0}, \quad \sigma_{xy1}|_{y=0} = \sigma_{xy2}|_{y=0} \quad (96)$$

Wave superposition in the two materials gives:

$$u_1 = u^i + u^r, \quad \sigma_1 = \sigma^i + \sigma^r \quad (\text{material 1}) \quad (97)$$

$$u_2 = u^t, \quad \sigma_2 = \sigma^t \quad (\text{material 2}) \quad (98)$$

Substitution of Eqs. (97) and (98) into Eqs. (95) and (96) yields the interface conditions in terms of incident, reflected, and transmitted waves as

$$-u_x^r|_{y=0} + u_x^t|_{y=0} = u_x^i|_{y=0} \quad (99)$$

$$-u_y^r|_{y=0} + u_y^t|_{y=0} = u_y^i|_{y=0} \quad (100)$$

$$-\sigma_{yy}^r|_{y=0} + \sigma_{yy}^t|_{y=0} = \sigma_{yy}^i|_{y=0} \quad (101)$$

$$-\sigma_{xy}^r|_{y=0} + \sigma_{xy}^t|_{y=0} = \sigma_{xy}^i|_{y=0} \quad (102)$$

Note that, in preparation for finding the solution, Eqs. (99), (100), and (102) were arranged such that the unknowns (reflected and transmitted wave displacements and stresses) are separated from the knowns (i.e., incident wave displacements and stresses).

### Coherence Condition

Substitution of Eqs. (83), (84), and (85) into Eq. (99) yields, upon simplification by  $ie^{-i\omega t}$ ,

$$\begin{aligned} & -\xi_p^r \hat{\Phi}^r e^{i\xi_p^r x} - \eta_S^r \hat{H}_z^r e^{i\xi_S^r x} + \xi_p^t \hat{\Phi}^t e^{i\xi_p^t x} - \eta_S^t \hat{H}_z^t e^{i\xi_S^t x} \\ &= \xi_p^i \hat{\Phi}^i e^{i\xi_p^i x} - \eta_S^i \hat{H}_z^i e^{i\xi_S^i x}, \quad \forall x \in \mathbb{R} \end{aligned} \quad (103)$$

For Eq. (103) to hold at any value of  $x$ , the coefficients of  $x$  in the complex exponentials must have the same value, i.e.,

$$\xi_p^i = \xi_S^i = \xi_p^r = \xi_S^r = \xi_p^t = \xi_S^t = \xi \quad (104)$$

where  $\xi$  is the common wavenumber along the  $x$ -direction. Equation (104) is called the **coherence condition**. To understand its implications, substitute Eq. (71) into Eq. (104) and get

$$\begin{aligned} \gamma_{P1} \sin \theta_P^i &= \gamma_{S1} \sin \theta_S^i = \gamma_{P1} \sin \theta_P^r = \gamma_{S1} \sin \theta_S^r = \gamma_{P2} \sin \theta_P^t = \gamma_{S2} \sin \theta_S^t \\ &= \xi \end{aligned} \quad (105)$$

Thus, it is apparent that the coherence condition Eq. (104) establishes a relationship between wave angles and wavenumbers.

### Snell's Law for Ultrasonics

Substitution of  $\gamma = \omega/c$  into Eq. (105) and division by  $\omega$  yields

$$\begin{aligned} \frac{\sin \theta_P^i}{c_{P1}} &= \frac{\sin \theta_S^i}{c_{S1}} = \frac{\sin \theta_P^r}{c_{P1}} = \frac{\sin \theta_S^r}{c_{S1}} = \frac{\sin \theta_P^t}{c_{P2}} = \frac{\sin \theta_S^t}{c_{S2}} \\ &= \frac{\xi}{\omega} \quad (\text{Snell's law for ultrasonics}) \end{aligned} \quad (106)$$

Equation (106) is similar to the Snell's law (aka law of sines) used in optics to determine the refraction angles when a beam of light passes from one medium into another; for this reason, it is called Snell's law for ultrasonics.

### Coherence of Incident P + SV Excitation

The first and second terms in Eq. (106) describe the coherence condition applied to the incident P and SV waves, respectively. In order for this coherence condition to be satisfied, the sines of the angles of the incident P and SV waves must be proportional to the respective wavespeeds,

$$\frac{\sin \theta_P^i}{c_{P1}} = \frac{\sin \theta_S^i}{c_{S1}} \quad (107)$$

Equation (107) represents Snell's law applied to incident P and SV excitation waves. If the incident P and SV excitation waves are coherent, i.e., if Eq. (107) is satisfied, then they share a common wavenumber  $\xi$  along the  $x$ -axis. In this case, their effects will combine coherently such that the result will be a single set of reflected and transmitted P + SV waves as shown in Fig. 1. However, if the incident P and SV excitations do not satisfy Eq. (107), then their effect will not combine coherently, each will generate its own set of reflected and transmitted P + SV waves, and the result will be four reflected waves and four transmitted waves (two P waves and two SV waves in each set).

### General Solution

A general solution can be developed under the assumption that the incident P and SV waves are coherent, i.e., Eq. (107) is satisfied. Then, Eq. (104) can be used to replace all the  $x$ -direction wavenumbers by a common wavenumber  $\xi$ . Next, Eq. (106) is used to find the reflection and transmission angles  $\theta_P^r, \theta_S^r, \theta_P^t, \theta_S^t$  in terms of  $\xi$  and  $\omega$ , i.e.,

$$\begin{aligned} \theta_P^i &= \theta_P^r = \theta_{P1} \\ \theta_S^i &= \theta_S^r = \theta_{S1} \quad \text{where} \quad \theta_{P1} = \sin^{-1} \left( \frac{\xi}{\omega} c_{P1} \right), \quad \theta_{S1} = \sin^{-1} \left( \frac{\xi}{\omega} c_{S1} \right) \\ \theta_P^t &= \theta_{P2} \quad \theta_{P2} = \sin^{-1} \left( \frac{\xi}{\omega} c_{P2} \right), \quad \theta_{S2} = \sin^{-1} \left( \frac{\xi}{\omega} c_{S2} \right) \\ \theta_S^t &= \theta_{S2} \end{aligned} \quad (108)$$

The wavenumbers are now calculated as

$$\begin{aligned} \gamma_P^i &= \gamma_P^r = \gamma_{P1} = \omega / c_{P1}, \quad \gamma_S^i = \gamma_S^r = \gamma_{S1} = \omega / c_{S1} \\ \gamma_P^r &= \gamma_{P2} = \omega / c_{P2} \quad \gamma_S^r = \gamma_{S2} = \omega / c_{S2} \end{aligned} \quad (109)$$

$$\xi_P^i = \xi_S^i = \xi_P^r = \xi_S^r = \xi_P^t = \xi_S^t = \xi \quad (110)$$

$$\begin{aligned} \eta_P^i &= \eta_P^r = \eta_{P1} = \gamma_{P1} \cos \theta_{P1}, \quad \eta_S^i = \eta_S^r = \eta_{S1} = \gamma_{S1} \cos \theta_{S1} \\ \eta_P^r &= \eta_{P2} = \gamma_{P2} \cos \theta_{P2} \quad \eta_S^r = \eta_{S2} = \gamma_{S2} \cos \theta_{S2} \end{aligned} \quad (111)$$

Substitution these results into Eqs. (83), (84), (85), (86), (87), (88), (89), (90), (91), (92), (93), and (94) yields

$$u_x^i|_{y=0} = i\left\{\xi_{P1}\hat{\Phi}^i - \eta_{S1}\hat{H}_z^i\right\}, \quad u_x^r|_{y=0} = i\left\{\xi_{P1}\hat{\Phi}^r + \eta_{S1}\hat{H}_z^r\right\}, \quad u_x^t|_{y=0} = i\left\{\xi_{P2}\hat{\Phi}^t - \eta_{S2}\hat{H}_z^t\right\} \quad (112)$$

$$u_y^i|_{y=0} = i\left\{-\eta_{P1}\hat{\Phi}^i - \xi_{S1}\hat{H}_z^i\right\}, \quad u_y^r|_{y=0} = i\left\{\eta_{P1}\hat{\Phi}^r - \xi_{S1}\hat{H}_z^r\right\}, \quad u_y^t|_{y=0} = i\left\{-\eta_{P2}\hat{\Phi}^t - \xi_{S2}\hat{H}_z^t\right\} \quad (113)$$

$$\sigma_{yy}^i|_{y=0} = [2\mu_1\xi_{P1}^2 - (\lambda_1 + 2\mu_1)\gamma_{P1}^2]\hat{\Phi}^i - 2\mu_1\xi_{S1}\eta_{S1}\hat{H}_z^i \quad (114)$$

$$\sigma_{yy}^r|_{y=0} = [2\mu_1\xi_{P1}^2 - (\lambda_1 + 2\mu_1)\gamma_{P1}^2]\hat{\Phi}^r + 2\mu_1\xi_{S1}\eta_{S1}\hat{H}_z^r \quad (115)$$

$$\sigma_{yy}^t|_{y=0} = [2\mu_2\xi_{P2}^2 - (\lambda_2 + 2\mu_2)\gamma_{P2}^2]\hat{\Phi}^t - 2\mu_2\xi_{S2}\eta_{S2}\hat{H}_z^t \quad (116)$$

$$\sigma_{xy}^i|_{y=0} = 2\mu_1\xi_{P1}\eta_{P1}\hat{\Phi}^i + \mu_1(\xi_{S1}^2 - \eta_{S1}^2)\hat{H}_z^i \quad (117)$$

$$\sigma_{xy}^r|_{y=0} = -2\mu_1\xi_{P1}\eta_{P1}\hat{\Phi}^r + \mu_1(\xi_{S1}^2 - \eta_{S1}^2)\hat{H}_z^r \quad (118)$$

$$\sigma_{xy}^t|_{y=0} = 2\mu_2\xi_{P2}\eta_{P2}\hat{\Phi}^t + \mu_2(\xi_{S2}^2 - \eta_{S2}^2)\hat{H}_z^t \quad (119)$$

where the common factor  $e^{i(\xi x - \omega t)}$  is implied but not explicitly written for the sake of brevity.

### Interface Conditions in Terms of Potentials

Substitution of Eqs. (112), (113), (114), (115), (116), (117), (118), and (119) into Eqs. (99), (100), (101), and (102) and division by  $e^{i(\xi x - \omega t)}$  yields

$$-\left(i\xi_{P1}\hat{\Phi}^r + i\eta_{S1}\hat{H}_z^r\right) + i\xi_{P2}\hat{\Phi}^t - i\eta_{S2}\hat{H}_z^t = i\xi_{P1}\hat{\Phi}^i - i\eta_{S1}\hat{H}_z^i \quad (120)$$

$$-\left(i\eta_{P1}\hat{\Phi}^r - i\xi_{S1}\hat{H}_z^r\right) - i\eta_{P2}\hat{\Phi}^t - i\xi_{S2}\hat{H}_z^t = -i\eta_{P1}\hat{\Phi}^i - i\xi_{S1}\hat{H}_z^i \quad (121)$$

$$\begin{aligned} & -\left\{[2\mu_1\xi_{P1}^2 - (\lambda_1 + 2\mu_1)\gamma_{P1}^2]\hat{\Phi}^r + 2\mu_1\xi_{S1}\eta_{S1}\hat{H}_z^r\right\} + [2\mu_2\xi_{P2}^2 - (\lambda_2 + 2\mu_2)\gamma_{P2}^2]\hat{\Phi}^t \\ & - 2\mu_2\xi_{S2}\eta_{S2}\hat{H}_z^t = [2\mu_1\xi_{P1}^2 - (\lambda_1 + 2\mu_1)\gamma_{P1}^2]\hat{\Phi}^i - 2\mu_1\xi_{S1}\eta_{S1}\hat{H}_z^i \end{aligned} \quad (122)$$

$$\begin{aligned}
& - \left[ -2\mu_1 \xi_{P1} \eta_{P1} \hat{\Phi}^r + \mu_1 (\xi_{S1}^2 - \eta_{S1}^2) \hat{H}_z^r \right] + 2\mu_2 \xi_{P2} \eta_{P2} \hat{\Phi}^t + \mu_2 (\xi_{S2}^2 - \eta_{S2}^2) \hat{H}_z^t \\
& = 2\mu_1 \xi_{P1} \eta_{P1} \hat{\Phi}^i + \mu_1 (\xi_{S1}^2 - \eta_{S1}^2) \hat{H}_z^i
\end{aligned} \tag{123}$$

Upon rearrangement and simplification, Eqs. (120), (121), (122), and (123) can be cast in matrix form, as

$$\mathbf{U}\mathbf{x} = \mathbf{q} \tag{124}$$

where

$$\mathbf{U} = \begin{bmatrix} -\xi_{P1} & -\eta_{S1} & \xi_{P2} & -\eta_{S2} \\ -\eta_{P1} & \xi_{S1} & -\eta_{P2} & -\xi_{S2} \\ -\mu_1 (\xi_{S1}^2 - \eta_{S1}^2) & -2\mu_1 \xi_{S1} \eta_{S1} & \mu_2 (\xi_{S2}^2 - \eta_{S2}^2) & -2\mu_2 \xi_{S2} \eta_{S2} \\ 2\mu_1 \xi_{P1} \eta_{P1} & -\mu_1 (\xi_{S1}^2 - \eta_{S1}^2) & 2\mu_2 \xi_{P2} \eta_{P2} & \mu_2 (\xi_{S2}^2 - \eta_{S2}^2) \end{bmatrix} \tag{125}$$

$$\mathbf{x} = \begin{bmatrix} \hat{\Phi}^r \\ \hat{H}_z^r \\ \hat{\Phi}^i \\ \hat{H}_z^i \end{bmatrix} \quad \mathbf{q} = \begin{bmatrix} \xi_{P1} \\ -\eta_{P1} \\ \mu_1 (\xi_{S1}^2 - \eta_{S1}^2) \\ 2\mu_1 \xi_{P1} \eta_{P1} \end{bmatrix} \hat{\Phi}^i + \begin{bmatrix} -\eta_{S1} \\ -\xi_{S1} \\ -2\mu_1 \xi_{S1} \eta_{S1} \\ \mu_1 (\xi_{S1}^2 - \eta_{S1}^2) \end{bmatrix} \hat{H}_z^i \tag{126}$$

Note that the identity  $2\mu\xi_P^2 - (\lambda + 2\mu)\gamma_P^2 = \mu(\xi_S^2 - \eta_S^2)$  was used to get from Eqs. (120), (121), (122), and (123) to Eqs. (125) and (126).

### Interface Conditions in Terms of Wave Angles and Material Properties

Equations (125) and (126) can be expressed in terms of only the wave angles  $\theta_{P1}, \theta_{S1}, \theta_{P2}, \theta_{S2}$  and the material properties in the two media, i.e.,

$$\mathbf{U} = \begin{bmatrix} -1 & -\cot\theta_{S1} & 1 & -\cot\theta_{S2} \\ -\cot\theta_{P1} & 1 & -\cot\theta_{P2} & -1 \\ -\mu_1(1 - \cot^2\theta_{S1}) & -2\mu_1 \cot\theta_{S1} & \mu_2(1 - \cot^2\theta_{S2}) & -2\mu_2 \cot\theta_{S2} \\ 2\mu_1 \cot\theta_{P1} & -\mu_1(1 - \cot^2\theta_{S1}) & 2\mu_2 \cot\theta_{P2} & \mu_2(1 - \cot^2\theta_{S2}) \end{bmatrix} \tag{127}$$

$$\mathbf{q} = \begin{bmatrix} 1 \\ -\cot\theta_{P1} \\ \mu_1(1 - \cot^2\theta_{S1}) \\ 2\mu_1 \cot\theta_{P1} \end{bmatrix} \hat{\Phi}^i + \begin{bmatrix} -\cot\theta_{S1} \\ -1 \\ -2\mu_1 \cot\theta_{S1} \\ \mu_1(1 - \cot^2\theta_{S1}) \end{bmatrix} \hat{H}_z^i \tag{128}$$

Note that use was made of Eqs. (110) and (111) as well as of the trigonometric relations

$$\frac{\eta}{\xi} = \frac{\gamma \cos \theta}{\gamma \sin \theta} = \cot \theta, \quad 1 - \frac{\eta^2}{\xi^2} = 1 - \frac{\cos^2 \theta}{\sin^2 \theta} = 1 - \cot^2 \theta \tag{129}$$

## Improve Numerical Conditioning by Normalization

The system (127) and (128) is badly conditioned because the first and second rows of matrix  $\mathbf{U}$  are much smaller than the following two rows, which would produce numerical difficulties during the solution process. The numerical conditioning can be improved by normalization of rows 3 and 4 to bring the elements of  $\mathbf{U}$  to the same order of magnitude, e.g.,

$$\mathbf{U} = \begin{bmatrix} -1 & -\cot \theta_{S1} & 1 & -\cot \theta_{S2} \\ -\cot \theta_{P1} & 1 & -\cot \theta_{P2} & -1 \\ -(1 - \cot^2 \theta_{S1}) & -2 \cot \theta_{S1} & \frac{\mu_2}{\mu_1} (1 - \cot^2 \theta_{S2}) & -2 \frac{\mu_2}{\mu_1} \cot \theta_{S2} \\ 2 \cot \theta_{P1} & -(1 - \cot^2 \theta_{S1}) & 2 \frac{\mu_2}{\mu_1} \cot \theta_{P2} & \frac{\mu_2}{\mu_1} (1 - \cot^2 \theta_{S2}) \end{bmatrix} \quad (130)$$

$$\mathbf{q} = \begin{bmatrix} 1 \\ -\cot \theta_{P1} \\ 1 - \cot^2 \theta_{S1} \\ 2 \cot \theta_{P1} \end{bmatrix} \hat{\Phi}^i + \begin{bmatrix} -\cot \theta_{S1} \\ -1 \\ -2 \cot \theta_{S1} \\ 1 - \cot^2 \theta_{S1} \end{bmatrix} \hat{H}_z^i \quad (131)$$

The solution is calculated numerically as

$$\mathbf{x} = \mathbf{U}^{-1} \mathbf{q} \quad \begin{bmatrix} \hat{\Phi}^r \\ \hat{H}_z^r \\ \hat{\Phi}^t \\ \hat{H}_z^t \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad (132)$$

Equation (132) provides the numerical values of the unknown potentials  $\hat{\Phi}^r, \hat{H}_z^r, \hat{\Phi}^t, \hat{H}_z^t$ .

## P-wave Excitation

If the incident excitation contains only the P-wave impinging at an incident angle  $\theta_P^i$ , then the solution can be deduced from the general solution by making  $H_z^i \equiv 0$ .

## Wave Angles for P-wave Excitation

The wave angles in the two materials take the form

$$\theta_{P1} = \theta_P^i \rightarrow \begin{aligned} \theta_P^r &= \theta_{P1} & \theta_{S1} &= \sin^{-1} \left( \frac{c_{S1}}{c_{P1}} \sin \theta_{P1} \right) \\ \theta_S^r &= \theta_{S1} & \text{where} & \\ \theta_P^t &= \theta_{P2} & \theta_{P2} &= \sin^{-1} \left( \frac{c_{P2}}{c_{P1}} \sin \theta_{P1} \right) \\ \theta_S^t &= \theta_{S2} & \theta_{S2} &= \sin^{-1} \left( \frac{c_{S2}}{c_{P1}} \sin \theta_{P1} \right) \end{aligned} \quad (133)$$

### Solution for P-wave Excitation

The governing equation is obtained from Eqs. (130) and (131) by setting  $H_z^i \equiv 0$ , i.e.,

$$\mathbf{U} = \begin{bmatrix} -1 & -\cot \theta_{S1} & 1 & -\cot \theta_{S2} \\ -\cot \theta_{P1} & 1 & -\cot \theta_{P2} & -1 \\ -(1 - \cot^2 \theta_{S1}) & -2 \cot \theta_{S1} & \frac{\mu_2}{\mu_1} (1 - \cot^2 \theta_{S2}) & -2 \frac{\mu_2}{\mu_1} \cot \theta_{S2} \\ 2 \cot \theta_{P1} & -(1 - \cot^2 \theta_{S1}) & 2 \frac{\mu_2}{\mu_1} \cot \theta_{P2} & \frac{\mu_2}{\mu_1} (1 - \cot^2 \theta_{S2}) \end{bmatrix} \quad (134)$$

$$\mathbf{x} = \begin{bmatrix} \hat{\Phi}^r \\ \hat{H}^r \\ \hat{\Phi}^i \\ \hat{H}_z^i \end{bmatrix} \quad \mathbf{q} = \begin{bmatrix} 1 \\ -\cot \theta_{P1} \\ 1 - \cot^2 \theta_{S1} \\ 2 \cot \theta_{P1} \end{bmatrix} \hat{\Phi}^i \quad (135)$$

The solution is calculated numerically as

$$\mathbf{x} = \mathbf{U}^{-1} \mathbf{q} \quad \begin{bmatrix} \hat{\Phi}^r \\ \hat{H}^r \\ \hat{\Phi}^i \\ \hat{H}_z^i \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad (136)$$

Note that although the excitation did not have an incident SV-wave, the solution contains both P-wave and SV-wave reflections and transmissions. This phenomenon is called *mode conversion* at the interface.

### Critical Angle for P-wave Excitation

The use of Eq. (133) to calculate the reflected and transmitted angles using the function  $\sin^{-1}$  is only possible if the argument of the  $\sin^{-1}$  is not greater than 1. As we change the incident P-wave angle  $\theta_{P1}$ , then the argument of  $\sin^{-1}$  functions in Eq. (133) also changes until a critical condition is reached when the argument equals to 1 corresponding to a transmission angle of  $90^\circ$ . Beyond this point, the  $\sin^{-1}$  function will no longer give a solution and the wave will no longer be transmitted in the 2<sup>nd</sup> material. Two critical angles may exist, one for the transmitted P-wave  $\theta_{P1}^{P2_{cr}}$  and the other for the transmitted S-wave  $\theta_{P1}^{S2_{cr}}$ , i.e.,

$$\begin{aligned} \theta_{P1}^{P2_{cr}} &= \sin^{-1}(c_{P1}/c_{P2}) \\ \theta_{P1}^{S2_{cr}} &= \sin^{-1}(c_{P1}/c_{S2}) \end{aligned} \quad (137)$$

since P-speed is always greater than the S-speed, we have

$$\theta_{P1}^{P2_{cr}} < \theta_{P1}^{S2_{cr}} \quad (138)$$

<i>Incident SV-wave</i>	$\theta_{P1} = 0^\circ$	$\theta_{P1}^{P2\text{ cr}}$	$\theta_{P1}^{S2\text{ cr}}$	$90^\circ$
<i>Transmitted wave</i>	$P$	$P + SV$	$SV$	<i>none</i>

**Fig. 4** Critical angles for incident P wave impinging obliquely on a bi-material interface

A depiction of the critical angles is given in Fig. 4. As the incident P-angle  $\theta_{P1}$  is experimentally increased from the vertical position  $\theta_{P1} = 0^\circ$  towards the horizontal position  $\theta_{P1} = 90^\circ$ , the P-critical angle  $\theta_{P1}^{P2\text{ cr}}$  occurs first and the transmitted P-wave becomes parallel with the interface ( $\theta_{P2} = 90^\circ$ ); this condition is known as *pressure grazing incidence*. Beyond this point, only the SV wave gets transmitted in the 2<sup>nd</sup> material. As the incident P-angle  $\theta_{P1}$  increases even further, the S-critical angle  $\theta_{P1}^{S2\text{ cr}}$  may be encountered (this second critical condition may or may not appear depending on the ratio  $c_{P1}/c_{S2}$  of the wavespeeds in the two materials). At this point, the transmitted SV-wave becomes parallel with the interface ( $\theta_{S2} = 90^\circ$ ); this condition is known as *shear grazing incidence*.

Beyond  $\theta_{P1}^{S2\text{ cr}}$ , there is no transmission into the 2<sup>nd</sup> material and the incident wave in the 1<sup>st</sup> material is just reflected back into the same material. This phenomenon is known as *total internal reflection*.

**Example 1** PMMA-aluminum interface with  $c_{P1} = 2,730 \text{ m/s}$ ,  $c_{S1} = 1,430 \text{ m/s}$ ,  $c_{P2} = 6,242 \text{ m/s}$ ,  $c_{S2} = 3,144 \text{ m/s}$  gives  $\theta_{P1}^{P2\text{ cr}} = \sin^{-1}(c_{P1}/c_{P2}) = 25.94^\circ$ ,  $\theta_{P1}^{S2\text{ cr}} = \sin^{-1}(c_{P1}/c_{S2}) = 60.26^\circ$ .

**Example 2** steel-aluminum interface with  $c_{P1} = 5,850 \text{ m/s}$ ,  $c_{S1} = 3,182 \text{ m/s}$ ,  $c_{P2} = 6,242 \text{ m/s}$ ,  $c_{S2} = 3,144 \text{ m/s}$ , gives  $\theta_{P1}^{P2\text{ cr}} = \sin^{-1}(c_{P1}/c_{P2}) = 69.49^\circ$  but no  $\theta_{P1}^{S2\text{ cr}} = \sin^{-1}(c_{P1}/c_{S2})$  because it comes out as an imaginary number. This is to be expected since the P wavespeed in material 1 is almost twice as larger as the S wavespeed in material 2.

## SV-wave Excitation

If the incident wave contains only the SV-wave impinging at an incident angle  $\theta_S^i$ , then the solution can be deduced from the general solution by making  $\Phi^i \equiv 0$ .

### Wave Angles for SV-wave Excitation

The wave angles in the two materials are obtained from Eq. (106) by solving in terms of  $\theta_S^i$ . Hence,

$$\begin{aligned} \theta_P^r &= \theta_{P1} & \theta_{P1} &= \sin^{-1}\left(\frac{c_{P1}}{c_{S1}} \sin \theta_{S1}\right) \\ \theta_S^r &= \theta_{S1} & \text{where} & \theta_{P2} = \sin^{-1}\left(\frac{c_{P2}}{c_{S1}} \sin \theta_{S1}\right) \\ \theta_P^t &= \theta_{P2} & & \theta_{S2} = \sin^{-1}\left(\frac{c_{S2}}{c_{S1}} \sin \theta_{S1}\right) \\ \theta_S^t &= \theta_{S2} & & \end{aligned} \quad (139)$$

### Solution for SV-wave Excitation

The governing equation is obtained from Eqs. (130) and (131) by setting  $\Phi^i \equiv 0$ , i.e.,

$$\mathbf{U} = \begin{bmatrix} -1 & -\cot \theta_{S1} & 1 & -\cot \theta_{S2} \\ -\cot \theta_{P1} & 1 & -\cot \theta_{P2} & -1 \\ -(1 - \cot^2 \theta_{S1}) & -2 \cot \theta_{S1} & \frac{\mu_2}{\mu_1} (1 - \cot^2 \theta_{S2}) & -2 \frac{\mu_2}{\mu_1} \cot \theta_{S2} \\ 2 \cot \theta_{P1} & -(1 - \cot^2 \theta_{S1}) & 2 \frac{\mu_2}{\mu_1} \cot \theta_{P2} & \frac{\mu_2}{\mu_1} (1 - \cot^2 \theta_{S2}) \end{bmatrix} \quad (140)$$

$$\mathbf{x} = \begin{bmatrix} \hat{\Phi}^r \\ \hat{H}^r \\ \hat{\Phi}^i \\ \hat{H}_z^t \end{bmatrix} \quad \mathbf{q} = \begin{bmatrix} -\cot \theta_{S1} \\ -1 \\ -2 \cot \theta_{S1} \\ 1 - \cot^2 \theta_{S1} \end{bmatrix} \hat{H}_z^i \quad (141)$$

The solution is calculated numerically as

$$\mathbf{x} = \mathbf{U}^{-1} \mathbf{q} \quad \begin{bmatrix} \hat{\Phi}^r \\ \hat{H}^r \\ \hat{\Phi}^i \\ \hat{H}_z^t \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad (142)$$

Note that although the excitation did not have an incident P-wave, the solution contains both P-wave and SV-wave reflections and transmissions. This phenomenon is called *mode conversion* at the interface.

### Critical Angle for SV-wave Excitation

The use of Eq. (139) to calculate the reflected and transmitted angles using the function  $\sin^{-1}$  is only possible if the argument of the  $\sin^{-1}$  is not greater than 1. As we change the incident SV-wave angle  $\theta_{S1}$ , then the argument of  $\sin^{-1}$  functions in Eq. (139) also changes until a critical condition is reached when the argument equals to 1 corresponding to a transmission angle of  $90^\circ$ . Beyond this point, the  $\sin^{-1}$  function will no longer give a solution and the wave will no longer be transmitted in the 2<sup>nd</sup> material. This incident angle is called *critical angle*. Three critical angles may exist, one for the reflected P-wave  $\theta_{S1}^{P1_{cr}}$ , another one for the transmitted P-wave  $\theta_{S1}^{P2_{cr}}$  and the third one for transmitted S-wave  $\theta_{S1}^{S2_{cr}}$ , i.e.,

$$\begin{aligned}\theta_{S1}^{P1_{cr}} &= \sin^{-1}(c_{S1}/c_{P1}) \\ \theta_{S1}^{P2_{cr}} &= \sin^{-1}(c_{S1}/c_{P2}) \\ \theta_{S1}^{S2_{cr}} &= \sin^{-1}(c_{S1}/c_{S2})\end{aligned}\quad (143)$$

since P-speed is always greater than the S-speed, we have

$$\theta_{S1}^{P2_{cr}} < \theta_{S1}^{S2_{cr}} \quad (144)$$

The relationship between  $\theta_{S1}^{P1_{cr}}$  and  $\theta_{S1}^{P2_{cr}}$  depends on the relative properties of the two materials.

In practice, as the incident SV-angle  $\theta_{S1}$  is experimentally increased from the vertical position  $\theta_{S1} = 0^\circ$  towards the horizontal position  $\theta_{S1} = 90^\circ$ , the phenomena of *pressure grazing incidence*, *shear grazing incidence*, and *total internal reflection* will also occur as described for the case of P-wave excitation in section “[P-wave Excitation](#).” However, the sequence of events may be more complicated as discussed next.

**Case 1** If  $\theta_{S1}^{P1_{cr}} < \theta_{S1}^{P2_{cr}}$ , then one gets the sequence events shown below where the reflected P-wave disappears first, then the transmitted P-wave and finally the transmitted SV-wave.

**Example 1** aluminum-steel interface with  $c_{P1} = 6,242$  m/s,  $c_{S1} = 3,144$  m/s,  $c_{P2} = 5,850$  m/s,  $c_{S2} = 3,182$  m/s, gives  $\theta_{S1}^{P1_{cr}} = \sin^{-1}(c_{S1}/c_{P1}) = 30.25^\circ$ ,  $\theta_{S1}^{P2_{cr}} = \sin^{-1}(c_{S1}/c_{P2}) = 32.51^\circ$ ,  $\theta_{S1}^{S2_{cr}} = \sin^{-1}(c_{S1}/c_{S2}) = 81.21^\circ$ . This situation is depicted in Fig. 5.

**Case 2** If  $\theta_{S1}^{P2_{cr}} < \theta_{S1}^{P1_{cr}}$ , then one gets the sequence of events shown below where the transmitted P-wave disappears first, then the reflected P-wave, and finally the transmitted S-wave.

**Example 2** copper-steel interface with  $c_{P1} = 4,450$  m/s,  $c_{S1} = 2,191$  m/s,  $c_{P2} = 5,850$  m/s,  $c_{S2} = 3,182$  m/s, gives  $\theta_{S1}^{P2_{cr}} = \sin^{-1}(c_{S1}/c_{P2}) = 22.00^\circ$ ,  $\theta_{S1}^{P1_{cr}} = \sin^{-1}(c_{S1}/c_{P1}) = 29.50^\circ$ ,  $\theta_{S1}^{S2_{cr}} = \sin^{-1}(c_{S1}/c_{S2}) = 43.52^\circ$ . This situation is depicted in Fig. 6

**Case 3** In some cases, the highest critical angle,  $\theta_{S1}^{S2_{cr}}$ , may not exist, i.e., it has imaginary values.

**Example 3a** steel-copper interface with  $c_{P1} = 5,850$  m/s,  $c_{S1} = 3,182$  m/s,  $c_{P2} = 4,450$  m/s,  $c_{S2} = 2,191$  m/s, gives  $\theta_{S1}^{P1_{cr}} = \sin^{-1}(c_{S1}/c_{P1}) = 32.95^\circ$ ,  $\theta_{S1}^{P2_{cr}} = \sin^{-1}(c_{S1}/c_{P2}) = 45.64^\circ$ , but no  $\theta_{S1}^{S2_{cr}} = \sin^{-1}(c_{S1}/c_{S2})$  because it comes out as an imaginary number. This situation is depicted in Fig. 7.

**Example 3b** steel-aluminum interface with  $c_{P1} = 5,850$  m/s,  $c_{S1} = 3,182$  m/s,  $c_{P2} = 6,242$  m/s,  $c_{S2} = 3,144$  m/s, gives  $\theta_{S1}^{P2_{cr}} = \sin^{-1}(c_{S1}/c_{P2}) = 30.65^\circ$ ,  $\theta_{S1}^{P1_{cr}} =$

$\sin^{-1}(c_{S1}/c_{P1}) = 32.95^\circ$ , but no  $\theta_{S1}^{S2\text{cr}} = \sin^{-1}(c_{S1}/c_{S2})$  because it comes out as an imaginary number. This situation is depicted in Fig. 8.

Other possibilities may also exist depending on the relative values of the critical angles.

## Bulk Waves at the Interface Between Liquid and Solid Media

Liquids, e.g., water, deform freely to fill the available shape. This means that liquids cannot sustain any shear stresses. This concept, which is self-evident under quasi-static conditions, is extended to dynamic conditions (e.g., waves) under the inviscid assumption. In this section, we start by recalling the mathematical expressions describing the propagation of bulk waves in solids. Then, we continue with a discussion of the boundary conditions at the interface between liquid and solid media.

## Bulk Waves in Liquids

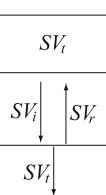
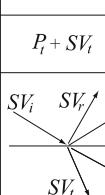
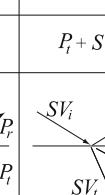
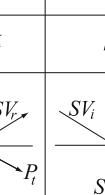
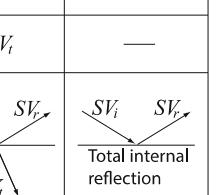
Inviscid liquids do not sustain shear stress. Hence, the deformation is purely dilatational. The constitutive relation in a liquid is

$$\sigma = \lambda \Delta \quad (\text{constitutive relation for liquids}) \quad (145)$$

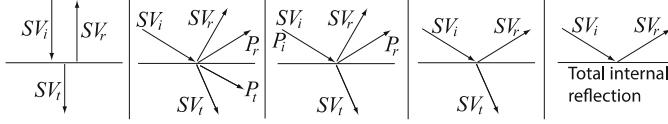
where  $\Delta$  is the dilation  $\Delta = \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz} = \vec{\nabla} \cdot \vec{u}$ . The 1<sup>st</sup> Lame constant  $\lambda$  is, in the case of liquids, the same as the bulk modulus, i.e.,  $\lambda = B$ .

## Wave Motion in Terms of Potentials

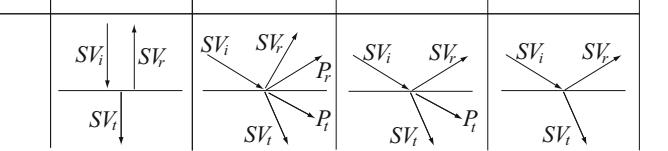
Bulk waves in a liquid can be described in terms of a single scalar potential  $\Phi$ , i.e.,

Incident SV-wave	$\theta_{s1}=0^\circ$	$\theta_{s1}^{P1\text{ cr}}$	$\theta_{s1}^{P2\text{ cr}}$	$\theta_{s1}^{S2\text{ cr}}$	$90^\circ$
Reflected	$SV_r$	$P_r + SV_r$	$SV_r$	$SV_r$	$SV_r$
Transmitted	$SV_t$	$P_t + SV_t$	$P_t + SV_t$	$SV_t$	—
					

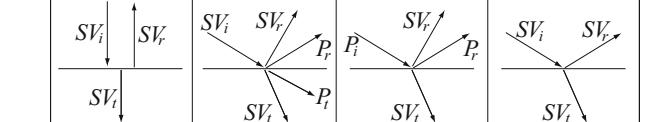
**Fig. 5** Critical angles for incident SV wave impinging obliquely on a bi-material interface. Case 1 when  $\theta_{S1}^{P1\text{cr}} < \theta_{S1}^{P2\text{cr}}$

<i>Incident SV-wave</i>	$\theta_{s1}=0^\circ$	$\theta_{s1}^{p2\ cr}$	$\theta_{s1}^{p1\ cr}$	$\theta_{s1}^{s2\ cr}$	$90^\circ$
<i>Reflected</i>	$SV_r$	$P_r + SV_r$	$P_r + SV_r$	$SV_r$	$SV_r$
<i>Transmitted</i>	$SV_t$	$P_t + SV_t$	$SV_t$	$SV_t$	—
					Total internal reflection

**Fig. 6** Critical angles for incident SV wave impinging obliquely on a bi-material interface. Case 2 when  $\theta_{s1}^{p2\ cr} < \theta_{s1}^{p1\ cr}$

<i>Incident SV-wave</i>	$\theta_{s1}=0^\circ$	$\theta_{s1}^{p1\ cr}$	$\theta_{s1}^{p2\ cr}$	$90^\circ$
<i>Reflected</i>	$SV_r$	$P_r + SV_r$	$SV_r$	$SV_r$
<i>Transmitted</i>	$SV_t$	$P_t + SV_t$	$P_t + SV_t$	$SV_t$
				

**Fig. 7** Critical angles for incident SV wave impinging obliquely on a bi-material interface. Case 3a when  $\theta_{s1}^{s2\ cr}$  does not exist and  $\theta_{s1}^{p1\ cr} < \theta_{s1}^{p2\ cr}$

<i>Incident SV-wave</i>	$\theta_{s1}=0^\circ$	$\theta_{s1}^{p2\ cr}$	$\theta_{s1}^{p1\ cr}$	$90^\circ$
<i>Reflected</i>	$SV_r$	$P_r + SV_r$	$P_r + SV_r$	$SV_r$
<i>Transmitted</i>	$SV_t$	$P_t + SV_t$	$SV_t$	$SV_t$
				

**Fig. 8** Critical angles for incident SV wave impinging obliquely on a bi-material interface. Case 3b when  $\theta_{s1}^{s2\ cr}$  does not exist and  $\theta_{s1}^{p1\ cr} > \theta_{s1}^{p2\ cr}$

$$\vec{u} = \vec{\nabla} \Phi \quad (\text{displacement potential in liquids}) \quad (146)$$

The displacement potential satisfies the wave equation, i.e.,

$$c^2 \nabla^2 \Phi = \ddot{\Phi} \quad (147)$$

where  $c^2 = \lambda/\rho$  is the wavespeed in the liquid. Note that the waves in a liquid are purely dilatational, i.e., P-waves.

For harmonic plane waves propagating in the direction  $\vec{n} = n_x \vec{i} + n_y \vec{j} + n_z \vec{k}$ , the displacement potential takes the form

$$\Phi(\vec{r}, t) = \hat{\Phi} e^{i(\vec{\gamma} \cdot \vec{r} - \omega t)} = \hat{\Phi} e^{i(\xi x + \eta y + \zeta z - \omega t)} \quad (\text{harmonic waves}) \quad (148)$$

where  $\vec{\gamma} = \gamma \vec{n}$ ,  $\vec{\gamma} = \xi \vec{i} + \eta \vec{j} + \zeta \vec{k}$ ,  $\gamma = \omega/c$ ,  $\xi = \gamma n_x$ ,  $\eta = \gamma n_y$ ,  $\zeta = \gamma n_z$ .

Substitution of Eq. (148) into Eq. (146) yields

$$\vec{u} = i \vec{\gamma} \Phi = i(\xi \vec{i} + \eta \vec{j} + \zeta \vec{k}) \Phi \quad (149)$$

or, by components,

$$\begin{aligned} u_x &= i\xi \Phi \\ u_y &= i\eta \Phi \\ u_z &= i\zeta \Phi \end{aligned} \quad (150)$$

## Stresses in Terms of Potentials

The stress tensor in liquids can be defined as a diagonal tensor of equal components, i.e.,

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} = \begin{bmatrix} \sigma & & \\ & \sigma & \\ & & \sigma \end{bmatrix} \quad (151)$$

Note that the stress tensor has only direct stresses since no shear stress may exist in a liquid medium. All the direct stresses are equal and their value does not depend on orientation, i.e.,  $\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = \sigma$ . Substitution of Eq. (146) into Eq. (145) yields the stresses in terms of potentials, i.e.,

$$\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = \sigma = \lambda \nabla^2 \Phi \quad (152)$$

For harmonic waves, Eq. (152) becomes

$$\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = \sigma = -\gamma^2 \lambda \Phi \quad (153)$$

## Interface Between Liquid and Solid Media

In liquids, only P-waves can propagate, whereas in solids both P and S waves may propagate. Hence, after setting up the general conditions at the interface between liquid and solid media, one would have to analyze separately the liquid-solid (LS) and the solid-liquid (SL) interfaces because the former can only have P-wave excitation, whereas the latter may have either P-wave or S-wave excitation, or both.

### Boundary Conditions

The boundary conditions at the interface between liquid and solid media are:

- Continuity of normal displacements
- Balance of normal stresses
- Vanishment of tangential stresses

One notices that only three interface conditions exist at a liquid-solid interface which is different from the situation at the solid-solid interface which has four interface conditions. This means that only three unknowns can be resolved at the interface between liquid and solid media. The equations describing the interface conditions between liquid and solid media are:

$$u_{y1}|_{y=0} = u_{y2}|_{y=0} \quad (\text{continuity of normal displacements}) \quad (154)$$

$$\sigma_{yy1}|_{y=0} = \sigma_{yy2}|_{y=0} \quad (\text{balance of normal stresses}) \quad (155)$$

$$\sigma_{xy}|_{y=0} = 0 \quad (\text{vanishment tangential stresses}) \quad (156)$$

Wave superposition in the two media yields:

$$u_1 = u^i + u^r, \quad \sigma_1 = \sigma^i + \sigma^r \quad (\text{in medium 1}) \quad (157)$$

$$u_2 = u^t, \quad \sigma_2 = \sigma^t \quad (\text{in medium 2}) \quad (158)$$

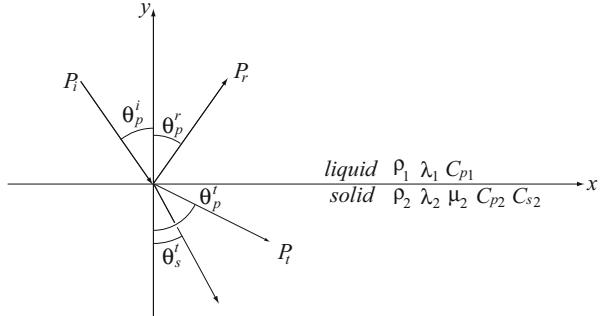
### LS and SL Interface Cases

In analyzing oblique waves impinging at the interface between liquid and solid media, one encounters two distinct situations:

1. The liquid-solid (LS) case: incident oblique P wave traveling in liquid impinges on a solid interface
2. The solid-liquid (SL) case: incident oblique P and/or SV wave traveling in solid impinges on the liquid interface

In the LS case, the result is a reflected P wave in the solid and two transmitted waves in the solid, P+SV. In the SL case, the situation is more complicated. If the

**Fig. 9** Oblique waves at a liquid-solid (LS) interface



incident is either P or SV, then one has two waves reflected in the solid, P+SV, and one wave transmitted in the liquid. If both P and SV are incident and they are not coherent at the interface, then one gets an overlap of cases with four waves reflected in the solid (P+SV due to incident P and P+SV due to incident SV) and two P waves transmitted into the liquid, one for the incident P and the incident SV. However, if the incident P and SV are coherent at the interface, then only two waves, P+SV, are reflected in the solid and one wave, P, is transmitted in the liquid. In the next section, the discussion will start with the simpler case of the LS interface. Then, the more complicated SL interface will be considered in a subsequent section.

## Bulk Waves at Liquid-Solid Interface

Consider a liquid-solid (LS) interface between materials with properties  $\rho_1, \lambda_1, c_{P1}$  and  $\rho_2, \lambda_2, \mu_2, c_{P2}, c_{S2}$ . An incident P wave impinges obliquely on the interface at angle  $\theta_p^i$  (Fig. 9). The potentials for the incident, reflected, and transmitted waves are

$$\Phi^i = \hat{\Phi}^i e^{i(\xi_p^i x - \eta_p^i y - \omega t)} \quad (\text{incident wave in the fluid}) \quad (159)$$

$$\Phi^r = \hat{\Phi}^r e^{i(\xi_p^r x + \eta_p^r y - \omega t)} \quad (\text{reflected wave in the fluid}) \quad (160)$$

$$\begin{aligned} \Phi^t &= \hat{\Phi}^t e^{i(\xi_p^t x - \eta_p^t y - \omega t)}, \\ H_z^t &= \hat{H}_z^t e^{i(\xi_s^t x - \eta_s^t y - \omega t)} \quad (\text{transmitted waves in the solid}) \end{aligned} \quad (161)$$

## Displacement and Stresses at the LS Interface

Substitution of Eqs. (159), (160) into Eqs. (149), (153) and of Eq. (161) into Eqs. (56), (66) followed by imposing  $y = 0$  yields the displacements and stresses at the interface as

$$u_y^i \Big|_{y=0} = -i\eta_P^i \hat{\Phi}^i e^{i(\xi_p^i x - \omega t)}, \quad u_y^r \Big|_{y=0} = i\eta_P^r \hat{\Phi}^r e^{i(\xi_p^r x - \omega t)} \quad (162)$$

$$\left. u_y^t \right|_{y=0} = -i\eta_P^t \hat{\Phi}^t e^{i(\xi_P^t x - \omega t)} - i\xi_S^t \hat{H}_z^t e^{i(\xi_S^t x - \omega t)} \quad (163)$$

$$\left. \sigma_{yy}^i \right|_{y=0} = -\lambda_1 \gamma_{P1}^2 \hat{\Phi}^i e^{i(\xi_P^i x - \omega t)}, \quad \left. \sigma_{yy}^r \right|_{y=0} = -\lambda_1 \gamma_{P1}^2 \hat{\Phi}^r e^{i(\xi_P^r x - \omega t)} \quad (164)$$

$$\left. \sigma_{yy}^t \right|_{y=0} = \left[ 2\mu_2 (\xi_P^t)^2 - (\lambda_2 + 2\mu_2) \gamma_{P2}^2 \right] \hat{\Phi}^t e^{i(\xi_P^t x - \omega t)} - 2\mu_2 \xi_S^t \eta_S^t \hat{H}_z^t e^{i(\xi_S^t x - \omega t)} \quad (165)$$

$$\left. \sigma_{xy}^i \right|_{y=0} = 0, \quad \left. \sigma_{xy}^r \right|_{y=0} = 0, \quad \left. \sigma_{xy}^t \right|_{y=0} = 2\mu_2 \xi_P^t \eta_P^t \hat{\Phi}^t e^{i(\xi_P^t x - \omega t)} + \mu_2 \left[ (\xi_S^t)^2 - (\eta_S^t)^2 \right] \hat{H}_z^t e^{i(\xi_S^t x - \omega t)} \quad (166)$$

### Interface Conditions in Terms of Incident, Reflected, and Transmitted Waves

In terms of incident, reflected, and transmitted waves, the interface conditions are

$$\left. -u_y^r \right|_{y=0} + \left. u_y^t \right|_{y=0} = \left. u_y^i \right|_{y=0} \quad (167)$$

$$\left. -\sigma_{yy}^r \right|_{y=0} + \left. \sigma_{yy}^t \right|_{y=0} = \left. \sigma_{yy}^i \right|_{y=0} \quad (168)$$

$$\left. \sigma_{xy}^t \right|_{y=0} = 0 \quad (169)$$

### Coherence Condition at SL Interface

Substitution of Eqs. (162), (163) into Eq. (167) yields, after simplification by  $ie^{-i\omega t}$ ,

$$-i\eta_P^r \hat{\Phi}^r e^{i\xi_P^r x} - i\eta_P^t \hat{\Phi}^t e^{i\xi_P^t x} - i\xi_S^t \hat{H}_z^t e^{i\xi_S^t x} = -i\eta_P^i \hat{\Phi}^i e^{i\xi_P^i x} \quad (170)$$

For Eq. (170) to hold at any value of  $x$ , the coefficients of  $x$  in the complex exponentials must have the same value, i.e.,

$$\xi_P^i = \xi_P^r = \xi_P^t = \xi_S^t = \xi \quad (171)$$

Equation (171) is the coherence condition which leads to Snell's law at the LS interface, i.e.,

$$\frac{\sin \theta_P^i}{c_{P1}} = \frac{\sin \theta_P^r}{c_{P1}} = \frac{\sin \theta_P^t}{c_{P2}} = \frac{\sin \theta_S^t}{c_{S2}} \quad \left( \text{Snell's law at LS interface} \right) \quad (172)$$

Hence,

$$\begin{aligned} \theta_{P1} &= \theta_P^i & \rightarrow & \theta_P^r = \theta_{P1} & \theta_{P2} &= \sin^{-1} \left( \frac{c_{P2}}{c_{P1}} \sin \theta_{P1} \right) \\ & \theta_P^t = \theta_{P2} & \text{where} & & \theta_{S2} &= \sin^{-1} \left( \frac{c_{S2}}{c_{P1}} \sin \theta_{P1} \right) \end{aligned} \quad (173)$$

The wavenumbers are now calculated as

$$\gamma_P^i = \gamma_P^r = \gamma_{P1} = \omega/c_{P1}, \quad \gamma_P^t = \gamma_{P2} = \omega/c_{P2}, \quad \gamma_S^t = \gamma_{S2} = \omega/c_{S2} \quad (174)$$

$$\xi_P^i = \xi_P^r = \xi_P^t = \xi_S^t = \xi \quad (175)$$

$$\eta_P^i = \eta_P^r = \eta_{P1} = \gamma_{P1} \cos \theta_{P1}, \quad \eta_P^t = \eta_{P2} = \gamma_{P2} \cos \theta_{P2}, \quad \eta_S^t = \eta_{S2} \\ = \gamma_{S2} \cos \theta_{S2} \quad (176)$$

### Interface Conditions in Terms of Potentials

Following the same procedure as for the case of solid-solid interface, one arrives at the following interface conditions in terms of potentials:

$$-\eta_{P1} \hat{\Phi}^r - \eta_{P2} \hat{\Phi}^t - \xi_{S2} \hat{H}_z^t = -\eta_{P1} \hat{\Phi}^i \quad (177)$$

$$\lambda_1 \gamma_{P1}^2 \hat{\Phi}^r + [2\mu_2 \xi_{P2}^2 - (\lambda_2 + 2\mu_2) \gamma_{P2}^2] \hat{\Phi}^t - 2\mu_2 \xi_{S2} \eta_{S2} \hat{H}_z^t = -\lambda_1 \gamma_{P1}^2 \hat{\Phi}^i \quad (178)$$

$$2\mu_2 \xi_{P2} \eta_{P2} \hat{\Phi}^t + \mu_2 (\xi_{S2}^2 - \eta_{S2}^2) \hat{H}_z^t = 0 \quad (179)$$

Equations (177), (178), and (179) can be cast in matrix form as

$$\mathbf{U}\mathbf{x} = \mathbf{q} \quad (180)$$

After simplification and normalization, the matrices  $\mathbf{U}$  and  $\mathbf{q}$  are found to be

$$\mathbf{U} = \begin{bmatrix} -\cot \theta_{P1} & -\cot \theta_{P2} & -1 \\ \frac{\lambda_1}{\mu_2} \csc^2 \theta_{P1} & 1 - \cot^2 \theta_{S2} & -2 \cot \theta_{S2} \\ 0 & 2 \cot \theta_{S2} & 1 - \cot^2 \theta_{S2} \end{bmatrix}, \quad (181)$$

$$\mathbf{q} = \begin{bmatrix} -\eta_{P1} \\ -\frac{\lambda_1}{\mu_2} \csc^2 \theta_{P1} \\ 0 \end{bmatrix} \hat{\Phi}^i$$

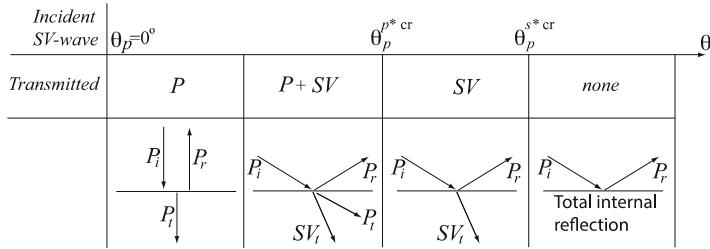
The solution is calculated numerically as

$$\mathbf{x} = \mathbf{U}^{-1} \mathbf{q} \quad \begin{bmatrix} \hat{\Phi}^r \\ \hat{\Phi}^t \\ \hat{H}_z^t \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (182)$$

Equation (182) provides numerical values for the unknown potentials  $\hat{\Phi}^r, \hat{\Phi}^t, \hat{H}_z^t$ . Note that although the excitation did not have an incident SV-wave, the transmission contains both P and SV waves. This phenomenon is called *mode conversion* at the LS interface.

### Critical Angles at Liquid Solid Interface

The use of Eq. (173) to calculate the reflected and transmitted angles using the function  $\sin^{-1}$  is only possible if the argument of the  $\sin^{-1}$  is not greater than 1. As



**Fig. 10** Critical angles for incident P wave impinging obliquely on liquid-solid (LS) interface

we change the incident P-wave angle  $\theta_{P1}$ , then the argument of  $\sin^{-1}$  functions in Eq. (173) also changes until a critical condition is reached when the argument equals to 1 corresponding to a transmission angle of  $90^\circ$ . Beyond this point, the  $\sin^{-1}$  function will no longer give a solution and the wave will no longer be transmitted in the 2<sup>nd</sup> material. Two critical angles may exist, one for the transmitted P-wave  $\theta_{P1}^{P2_{cr}}$  and the other for the transmitted S-wave  $\theta_{P1}^{S2_{cr}}$ , i.e.,

$$\theta_{P1}^{P2_{cr}} = \sin^{-1}(c_{P1}/c_{P2}), \quad \theta_{P1}^{S2_{cr}} = \sin^{-1}(c_{P1}/c_{S2}) \quad (183)$$

since P-speed is always greater than the S-speed, we have

$$\theta_{P1}^{P2_{cr}} < \theta_{P1}^{S2_{cr}} \quad (184)$$

A depiction of the critical angles is given in Fig. 10. As the incident P-angle  $\theta_{P1}$  is experimentally increased from the vertical position  $\theta_{P1} = 0^\circ$  towards the horizontal position  $\theta_{P1} = 90^\circ$ , the P-critical angle occurs first and the transmitted P-wave becomes parallel with the interface ( $\theta_{P2} = 90^\circ$ ); this condition is known as *pressure grazing incidence*. Beyond this point, only the SV wave gets transmitted in the 2<sup>nd</sup> material. As the incident P-angle  $\theta_{P1}$  increases even further, the S-critical angle  $\theta_{P1}^{S2_{cr}}$  may be encountered (this second critical condition may or may not appear depending on the ratio  $c_{P1}/c_{S2}$  of the wavespeeds in the two materials). At this point, the transmitted SV-wave becomes parallel with the interface ( $\theta_{S2} = 90^\circ$ ); this condition is known as *shear grazing incidence*.

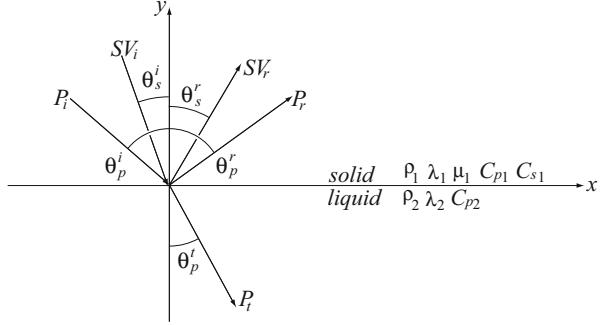
Beyond  $\theta_{P1}^{S2_{cr}}$ , there is no transmission into the 2<sup>nd</sup> material and the incident wave in the 1<sup>st</sup> material is just reflected back into the same material. This phenomenon is known as *total internal reflection*.

**Example** consider the liquid is water with  $c_{P1} = 1,500$  m/s, whereas the solid is aluminum with  $c_{P2} = 6,150$  m/s and  $c_{S2} = 3,100$  m/s. The critical angles are  $\theta_{P1}^{P2_{cr}} = \sin^{-1}(1,500/6,150) \approx 14^\circ$ ,  $\theta_{P1}^{S2_{cr}} = \sin^{-1}(1,500/3,100) \approx 29^\circ$ .

## Bulk Waves at Solid-Liquid Interface

Consider a solid-liquid (SL) interface between materials with properties  $\rho_1, \lambda_1, \mu_1, c_{P1}, c_{S2}$  and  $\rho_2, \lambda_2, c_{P2}$ . An incident P wave and/or an incident S wave impinge

**Fig. 11** Oblique waves at a solid-liquid (SL) interface



obliquely on the interface with angles  $\theta_p^i$ ,  $\theta_S^i$ , respectively (Fig. 11). The potentials for the incident, reflected, and transmitted waves are

$$\Phi^i = \hat{\Phi}^i e^{i(\xi_p^i x - \eta_p^i y - \omega t)}, \quad H_z^i = \hat{H}_z^i e^{i(\xi_s^i x - \eta_s^i y - \omega t)} \quad (\text{incident wave in the solid}) \quad (185)$$

$$\Phi^r = \hat{\Phi}^r e^{i(\xi_p^r x + \eta_p^r y - \omega t)}, \quad H_z^r = \hat{H}_z^r e^{i(\xi_s^r x + \eta_s^r y - \omega t)} \quad (\text{reflected wave in the solid}) \quad (186)$$

$$\Phi^t = \hat{\Phi}^t e^{i(\xi_p^t x - \eta_p^t y - \omega t)} \quad (\text{transmitted wave in the liquid}) \quad (187)$$

### Displacement and Stresses at the SL Interface

Substitution of Eqs. (185), (186) into Eqs. (56), (66) and of Eq. (187) into Eqs. (149), (153) followed by imposing  $y = 0$  yields the displacements and stresses at the interface as

$$u_y^i \Big|_{y=0} = -i\eta_p^i \hat{\Phi}^i e^{i(\xi_p^i x - \omega t)} - i\xi_s^i \hat{H}_z^i e^{i(\xi_s^i x - \omega t)}, \quad u_y^r \Big|_{y=0} = i\eta_p^r \hat{\Phi}^r e^{i(\xi_p^r x - \omega t)} - i\xi_s^r \hat{H}_z^r e^{i(\xi_s^r x - \omega t)} \quad (188)$$

$$u_y^t \Big|_{y=0} = -i\eta_p^t \hat{\Phi}^t e^{i(\xi_p^t x - \omega t)} \quad (189)$$

$$\sigma_{yy}^i \Big|_{y=0} = \left[ 2\mu_1 (\xi_p^i)^2 - (\lambda_1 + 2\mu_1) \gamma_{P1}^2 \right] \hat{\Phi}^i e^{i(\xi_p^i x - \omega t)} - 2\mu_1 \xi_s^i \eta_s^i \hat{H}_z^i e^{i(\xi_s^i x - \omega t)} \quad (190)$$

$$\sigma_{yy}^r \Big|_{y=0} = \left[ 2\mu_1 (\xi_p^r)^2 - (\lambda_1 + 2\mu_1) \gamma_{P1}^2 \right] \hat{\Phi}^r e^{i(\xi_p^r x - \omega t)} + 2\mu_1 \xi_s^r \eta_s^r \hat{H}_z^r e^{i(\xi_s^r x - \omega t)} \quad (191)$$

$$\sigma_{xy}^i \Big|_{y=0} = 2\mu_1 \xi_p^i \eta_p^i \hat{\Phi}^i e^{i(\xi_p^i x - \omega t)} + \mu_1 \left[ (\xi_s^i)^2 - (\eta_s^i)^2 \right] \hat{H}_z^i e^{i(\xi_s^i x - \omega t)} \quad (192)$$

$$\sigma_{xy}^r \Big|_{y=0} = -2\mu_1 \xi_p^r \eta_p^r \hat{\Phi}^r e^{i(\xi_p^r x - \omega t)} + \mu_1 \left[ (\xi_s^r)^2 - (\eta_s^r)^2 \right] \hat{H}_z^r e^{i(\xi_s^r x - \omega t)} \quad (193)$$

$$\sigma_{yy}^t \Big|_{y=0} = -\lambda_1 \gamma_{P2}^2 \hat{\Phi}^t e^{i(\xi_p^t x - \omega t)}, \quad \sigma_{xy}^t \Big|_{y=0} = 0 \quad (194)$$

### Interface Conditions in Terms of Incident, Reflected, and Transmitted Waves

In terms of incident, reflected, and transmitted waves, the interface conditions are

$$-u_y^r \Big|_{y=0} + u_y^t \Big|_{y=0} = u_y^i \Big|_{y=0} \quad (195)$$

$$-\sigma_{yy}^r \Big|_{y=0} + \sigma_{yy}^t \Big|_{y=0} = \sigma_{yy}^i \Big|_{y=0} \quad (196)$$

$$-\sigma_{xy}^r \Big|_{y=0} = \sigma_{xy}^i \Big|_{y=0} \quad (197)$$

### Coherence Condition at SL Interface

Substitution of Eqs. (188), (189) into Eq. (195) yields, after simplification by  $ie^{-i\omega t}$ ,

$$\eta_P^r \hat{\Phi}^r e^{i\xi_p^r x} - i\xi_S^r \hat{H}_z^r e^{i\xi_S^r x} - \eta_P^t \hat{\Phi}^t e^{i\xi_p^t x} = -\eta_P^i \hat{\Phi}^i e^{i\xi_p^i x} - \xi_S^i \hat{H}_z^i e^{i\xi_S^i x} \quad (198)$$

For Eq. (198) to hold at any value of  $x$ , the coefficients of  $x$  in the complex exponentials must have the same value, i.e.,

$$\xi_p^i = \xi_S^i = \xi_p^r = \xi_S^r = \xi_p^t = \xi \quad (199)$$

Equation (199) is the coherence condition which leads to Snell's law at the SL interface, i.e.,

$$\begin{aligned} \frac{\sin \theta_P^i}{c_{P1}} &= \frac{\sin \theta_S^i}{c_{S1}} = \frac{\sin \theta_P^r}{c_{P1}} = \frac{\sin \theta_S^r}{c_{S1}} = \frac{\sin \theta_P^t}{c_{P2}} \\ &= \frac{\xi}{\omega} \quad (\text{Snell's law at SL interface}) \end{aligned} \quad (200)$$

Hence,

$$\begin{array}{ll} \theta_P^i = \theta_{P1} & \theta_{P1} = \sin^{-1} \left( \frac{\xi}{\omega} c_{P1} \right) \\ \theta_S^i = \theta_{S1} & \\ \theta_P^r = \theta_{P1} & \text{where } \theta_{S1} = \sin^{-1} \left( \frac{\xi}{\omega} c_{S1} \right) \\ \theta_S^r = \theta_{S1} & \\ \theta_P^t = \theta_{P2} & \theta_{P2} = \sin^{-1} \left( \frac{\xi}{\omega} c_{P2} \right) \end{array} \quad (201)$$

The wavenumbers are now calculated as

$$\gamma_P^i = \gamma_P^r = \gamma_{P1} = \omega/c_{P1}, \quad \gamma_S^i = \gamma_S^r = \gamma_{S1} = \omega/c_{S1}, \quad \gamma_P^t = \gamma_{P2} = \omega/c_{P2} \quad (202)$$

$$\xi_P^i = \xi_S^i = \xi_P^r = \xi_S^r = \xi_P^t = \xi_S^t = \xi \quad (203)$$

$$\begin{aligned} \eta_P^i &= \eta_P^r = \eta_{P1} = \gamma_{P1} \cos \theta_{P1}, \quad \eta_S^i = \eta_S^r = \eta_{S1} = \gamma_{S1} \cos \theta_{S1}, \quad \eta_P^t = \eta_{P2} \\ &= \gamma_{P2} \cos \theta_{P2} \end{aligned} \quad (204)$$

### Interface Conditions in Terms of Potentials

Following the same procedure as for the case of solid-solid interface and assuming that the incident P and SV waves are coherent according to Eq. (201), one arrives at the following interface conditions in terms of potentials:

$$-\left(i\eta_{P1}\hat{\Phi}^r - i\xi_{S1}\hat{H}_z^r\right) - i\eta_{P2}\hat{\Phi}^t = -i\eta_{P1}\hat{\Phi}^i - i\xi_{S1}\hat{H}_z^i \quad (205)$$

$$\begin{aligned} &-\left\{[2\mu_1\xi_{P1}^2 - (\lambda_1 + 2\mu_1)\gamma_{P1}^2]\hat{\Phi}^r + 2\mu_1\xi_{S1}\eta_{S1}\hat{H}_z^r\right\} - \lambda_2\gamma_{P2}^2\hat{\Phi}^t \\ &= [2\mu_1\xi_{P1}^2 - (\lambda_1 + 2\mu_1)\gamma_{P1}^2]\hat{\Phi}^i - 2\mu_1\xi_{S1}\eta_{S1}\hat{H}_z^i \end{aligned} \quad (206)$$

$$2\mu_1\xi_{P1}\eta_{P1}\hat{\Phi}^r - \mu_1(\xi_{S1}^2 - \eta_{S1}^2)\hat{H}_z^r = 2\mu_1\xi_{P1}\eta_{P1}\hat{\Phi}^i + \mu_1(\xi_{S1}^2 - \eta_{S1}^2)\hat{H}_z^i \quad (207)$$

Equations (205), (206), and (207) can be cast in matrix form as

$$\mathbf{x} = \mathbf{U}^{-1}\mathbf{q} \quad (208)$$

After simplification and normalization, the matrices  $\mathbf{U}$  and  $\mathbf{q}$  are found to be

$$\mathbf{U} = \begin{bmatrix} -\cot \theta_{P1} & 1 & -\frac{\cot \theta_{P2}}{\mu_1} \\ (1 - \cot^2 \theta_{S1}) & -2 \cot \theta_{S1} & -\frac{\lambda_2}{\mu_1} \csc^2 \theta_{P2} \\ 2 \cot \theta_{S1} & -(1 - \cot^2 \theta_{S1}) & 0 \end{bmatrix} \quad (209)$$

$$\mathbf{q} = \begin{bmatrix} -\cot \theta_{P1} \\ 1 - \cot^2 \theta_{S1} \\ 2 \cot \theta_{P1} \end{bmatrix} \hat{\Phi}^i + \begin{bmatrix} -1 \\ -2 \cot \theta_{S1} \\ 1 - \cot^2 \theta_{S1} \end{bmatrix} \hat{H}_z^i \quad (210)$$

The solution is calculated numerically as

$$\mathbf{x} = \mathbf{U}^{-1}\mathbf{q}, \quad \begin{bmatrix} \hat{\Phi}^r \\ \hat{H}_z^r \\ \hat{\Phi}^i \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (211)$$

Equation (211) provides numerical values for the unknown potentials  $\hat{\Phi}^r, \hat{H}_z^r, \hat{\Phi}^i$ . Note that this solution assumes that the incident P and SV waves are coherent, i.e., their angles obey the constraints of Eq. (201). If this is not true, then one has to consider separately P-wave and SV-wave excitations, as shown next.

### P-wave Excitation at Solid-Liquid Interface

If the incident wave contains only the P-wave, then the solution can be deduced from the general solution by making  $H_z^i \equiv 0$ . Solution of Eq. (200) in terms of incident P-wave angle  $\theta_P^i$  yields the wave angles in the two materials as

$$\begin{aligned} \theta_{P1} = \theta_P^i &\rightarrow \theta_P^r = \theta_{P1} & \theta_{S1} = \sin^{-1} \left( \frac{c_{S1}}{c_{P1}} \sin \theta_{P1} \right) \\ &\theta_S^r = \theta_{S1} & \text{where} \\ &\theta_P^t = \theta_{P2} & \theta_{P2} = \sin^{-1} \left( \frac{c_{P2}}{c_{P1}} \sin \theta_{P1} \right) \end{aligned} \quad (212)$$

The  $\mathbf{q}$  term in Eq. (210) retains only the  $\hat{\Phi}^i$  part and Eqs. (209) and (210) become

$$\begin{aligned} \mathbf{U} &= \begin{bmatrix} -\cot \theta_{P1} & 1 & -\cot \theta_{P2} \\ (1 - \cot^2 \theta_{S1}) & -2 \cot \theta_{S1} & -\frac{\lambda_2}{\mu_1} \csc^2 \theta_{P2} \\ 2 \cot \theta_{S1} & -(1 - \cot^2 \theta_{S1}) & 0 \end{bmatrix}, \\ \mathbf{q} &= \begin{bmatrix} -\cot \theta_{P1} \\ 1 - \cot^2 \theta_{S1} \\ 2 \cot \theta_{P1} \end{bmatrix} \hat{\Phi}^i \end{aligned} \quad (213)$$

Substitution of Eq. (213) into Eq. (211) yields the solution

$$\mathbf{x} = \mathbf{U}^{-1} \mathbf{q}, \quad \begin{bmatrix} \hat{\Phi}^r \\ \hat{H}_z^r \\ \hat{\Phi}^t \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (214)$$

Note that although the excitation did not have an incident SV-wave, the solution contains both P-wave and SV-wave reflections. This phenomenon is called *mode conversion* at the interface.

The use of Eq. (212) to calculate the reflected and transmitted angles using the function  $\sin^{-1}$  is only possible if the argument of the  $\sin^{-1}$  is not greater than 1. As we change the incident P-wave angle  $\theta_{P1}$ , then the argument of  $\sin^{-1}$  functions in Eq. (212) also changes and the possibility of a critical condition for which the argument equals to 1 corresponding to a transmission angle of  $90^\circ$  may exist. Examination of Eq. (212) reveals that  $\theta_{P2}$  will always exist because liquid (e.g., water) wavespeed is always less than P-wavespeed in common materials, i.e.,  $c_{P2} < c_{P1}$ . Hence, no critical angles exist for oblique P-wave excitation at common solid-liquid interfaces.

### SV-wave Excitation at Solid-Liquid Interface

If the incident wave contains only the SV-wave, then the solution can be deduced from the general solution by making  $\Phi^i \equiv 0$ . Solution of Eq. (200) in terms of incident SV-wave angle  $\theta_S^i$  yields the wave angles in the two materials as

$$\begin{aligned} \theta_{S1} = \theta_S^i &\rightarrow \theta_P^r = \theta_{P1} & \theta_{P1} = \sin^{-1} \left( \frac{c_{P1}}{c_{S1}} \sin \theta_{S1} \right) \\ &\theta_S^r = \theta_{S1} & \text{where} \\ &\theta_P^t = \theta_{P2} & \theta_{P2} = \sin^{-1} \left( \frac{c_{P2}}{c_{S1}} \sin \theta_{S1} \right) \end{aligned} \quad (215)$$

The  $\mathbf{q}$  term in Eq. (210) retains only the  $\hat{H}_z^i$  part and Eqs. (209) and (210) become

$$\mathbf{U} = \begin{bmatrix} -\cot\theta_{P1} & 1 & -\cot\theta_{P2} \\ (1 - \cot^2\theta_{S1}) & -2\cot\theta_{S1} & -\frac{\lambda_2}{\mu_1}\csc^2\theta_{P2} \\ 2\cot\theta_{S1} & -(1 - \cot^2\theta_{S1}) & 0 \end{bmatrix}, \quad (216)$$

$$\mathbf{q} = \begin{bmatrix} -1 \\ -2\cot\theta_{S1} \\ 1 - \cot^2\theta_{S1} \end{bmatrix} \hat{H}_z^i$$

Substitution of Eq. (216) into Eq. (211) yields the solution

$$\mathbf{x} = \mathbf{U}^{-1}\mathbf{q}, \quad \begin{bmatrix} \hat{\Phi}^r \\ \hat{H}_z^r \\ \hat{\Phi}_t^i \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (217)$$

Note that although the excitation did not have an incident P-wave, the solution contains both P-wave and SV-wave reflections. This phenomenon is called *mode conversion* at the interface.

The use of Eq. (215) to calculate the reflected and transmitted angles using the function  $\sin^{-1}$  is only possible if the argument of the  $\sin^{-1}$  is not greater than 1. As we change the incident P-wave angle  $\theta_{P1}$ , then the argument of  $\sin^{-1}$  functions in Eq. (212) also changes and the possibility of a critical condition for which the argument equals to 1 corresponding to a transmission angle of  $90^\circ$  may exist. Two critical angles may exist: one for the reflected P-wave  $\theta_{S1}^{P1_{cr}}$ , another one for the transmitted P-wave  $\theta_{S1}^{P2_{cr}}$ , i.e.,

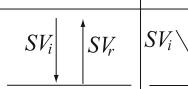
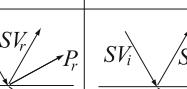
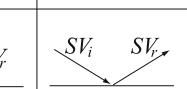
$$\begin{aligned} \theta_{S1}^{P1_{cr}} &= \sin^{-1}(c_{S1}/c_{P1}) \\ \theta_{S1}^{P2_{cr}} &= \sin^{-1}(c_{S1}/c_{P2}) \end{aligned} \quad (218)$$

**Case 1** both critical angles exist.

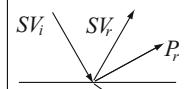
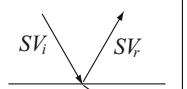
**Example 1** PMMA-water interface with  $c_{P1} = 2,730$  m/s,  $c_{S1} = 1,430$  m/s,  $c_{P2} = 1,500$  m/s, gives  $\theta_{S1}^{P1_{cr}} = \sin^{-1}(c_{S1}/c_{P1}) = 31.59^\circ$ ,  $\theta_{S1}^{P2_{cr}} = \sin^{-1}(c_{S1}/c_{P2}) = 72.43^\circ$ . A depiction of the critical angles is given in Fig. 12.

**Case 2** only one critical angle exists.

**Example 2** aluminum-water interface with  $c_{P1} = 6,242$  m/s,  $c_{S1} = 3,144$  m/s,  $c_{P2} = 1,500$  m/s, gives  $\theta_{S1}^{P1_{cr}} = \sin^{-1}(c_{S1}/c_{P1}) = 30.25^\circ$ , but no  $\theta_{S1}^{P2_{cr}} = \sin^{-1}(c_{S1}/c_{P2})$  because it comes out as an imaginary number. No total internal reflection exists in this case and a pressure wave will always propagate into the liquid medium. A depiction of the critical angles is given in Fig. 13.

<i>Incident SV-wave</i>	$\theta_{s1}=0^\circ$	$\theta_{s1}^{p1\ cr}$	$\theta_{s2}^{p2\ cr}$	$90^\circ$
<i>Reflected</i>	$SV_r$	$P_r + SV_r$	$SV_r$	$SV_r$
<i>Transmitted</i>		$P_t$	$P_t + SV_t$	$SV_t$
				

**Fig. 12** Critical angles for incident SV wave impinging obliquely on solid-liquid interface Case 1 when both critical angles exist

<i>Incident SV-wave</i>	$\theta_{s1}=0^\circ$	$\theta_{s1}^{p1\ cr}$	$90^\circ$
<i>Reflected</i>	$SV_r$	$P_r + SV_r$	$SV_r$
<i>Transmitted</i>		$P_t$	$P_t + SV_t$
			

**Fig. 13** Critical angles for incident SV wave impinging obliquely on solid-liquid interface. Case 2 when only one critical angle exists

## References

- Achenbach JD (1973) Wave propagation in elastic solids. Elsevier
- Graff KF (1991) Wave motion in elastic solids. Dover Publications, New York
- Kolsky H (1963) Stress waves in solids. Dover Publications, New York
- Krautkramer J, Krautkramer H (1990) Ultrasonic testing of materials. Springer
- Rose JL (1999) Ultrasonic waves in solid media. Cambridge University Press, New York
- Royer D, Dieulesaint E (2000) Elastic waves in solids. Springer



# Ultrasonic Wavefield Imaging

2

Jennifer E. Michaels

## Contents

Introduction .....	44
Acquiring and Visualizing Ultrasonic Wavefield Data .....	45
Analyzing Ultrasonic Wavefield Data .....	48
Phase and Group Velocities .....	48
Energy Imaging .....	50
Direct Arrival Imaging .....	51
Frequency-Wavenumber Analysis .....	53
Frequency-Wavenumber Filtering .....	56
Guided Waves in a Composite Panel .....	57
Guided Waves in a Bonded Plate .....	61
Bulk Waves in an Aluminum Plate .....	65
Additional Wavefield Methods .....	70
Summary .....	71
References .....	72

### Abstract

Ultrasonic wavefield imaging, or the acquisition and subsequent analysis of wave motion generated by a fixed source, is a powerful research tool that is also being considered for nondestructive evaluation (NDE). Many proposed wavefield-based NDE methods have come from the structural health monitoring (SHM) community and are based upon guided waves. If guided wave transducers are already embedded in or mounted on a structure as part of an SHM system, then wavefield-based inspection can potentially take place with very little required disassembly. Wavefield inspection methods may also be practical for stand-alone inspection because of the richness of the data and the noncontact nature of most

---

J. E. Michaels (✉)

School of Electrical and Computer Engineering, Georgia Institute of Technology,  
Atlanta, GA, USA

e-mail: [jemichaels@gatech.edu](mailto:jemichaels@gatech.edu)

implementations. In its simplest form, wavefield imaging is a very effective method for visualizing wave motion on the surface of a specimen. However, for it to become a practical NDE method, analysis methods must be applied to the wavefield data to yield quantitative information concerning the detection, localization, and characterization of damage. The focus of this chapter is to provide an introduction to the acquisition and analysis of wavefield data in the context of three case studies: impact damage in composites, bond evaluation for aluminum plates, and characterization of scattering from notched through-holes. References to additional methods and applications are given for the interested reader.

---

## Introduction

The intuitive value of ultrasonic wave imaging (UWI) can be readily appreciated by anyone who has watched water waves in a pond or lake resulting from a point-like disturbance. Both the nature of wave propagation and the interactions of the waves with obstacles are clearly visible. When applied to nondestructive evaluation (NDE), UWI provides a means to not only evaluate how waves propagate in a specimen of interest but also how they interact with both structural features and defects. Watching a “movie” of wave motion resulting from a stationary source of ultrasonic waves provides qualitative insight that can lead to improved inspection methods, and subsequent analysis of the wavefield data can provide quantitative information regarding both material properties and damage.

As is the case for conventional pulse-echo and through-transmission data acquisition with an automated scanning system, wavefield data are typically acquired by repetitively exciting the source and moving the receiver in a raster pattern over the accessible surface of the part. Unlike conventional ultrasonic methods, the source is stationary and only the receiver moves from point-to-point. The raster scanning process enables the complete wavefield to be acquired over a region of interest. Wavefield visualization occurs by displaying the data as a movie; that is, as a series of spatial frames or “snapshots” that update as time advances. In this manner one can view the waves propagating outward from the source and subsequently interacting with geometrical features and damage. Although various implementations of wavefield imaging have existed for many years (Wyatt 1972; Hall 1977; Nishizawa et al. 1997), the widespread availability of laser Doppler vibrometers that can acquire signals in the high kHz and low MHz range has led to a rapid expansion of research and development utilizing full wavefield data.

UWI is perhaps most suited to the measurement of guided waves since their propagation is guided by the accessible surface, which is also the measurement surface for wavefield acquisition. Guided waves are frequently used as an NDE screening method because of their ability to propagate long distances and remain sensitive to damage, but their long wavelengths relative to critical flaw sizes make damage characterization challenging. Wavefield imaging has greatly contributed to the understanding of how guided waves propagate and interact with damage such as fatigue cracks, impacts, and corrosion. Researchers in the area of structural health

monitoring (SHM), also referred to as *in situ* NDE, are proposing arrays of permanently attached transducers to continuously monitor critical structures using guided waves. If such an SHM system generates an alarm, a more thorough inspection will be required. Using one or more of the attached SHM transducers as sources of guided waves for UWI has thus been proposed as a follow-up NDE method that has the desirable characteristic of being noncontact (Michaels and Michaels 2006; Gannon et al. 2015).

Application of UWI to bulk waves is also possible, but the recorded data captures only the portion of the wavefield that interacts with the accessible surface, primarily by reflection but also via skimming waves. Nevertheless, it is possible to extract quantitative information from the wavefield regarding both propagation and scattering. Since conventional ultrasonic inspection methods have the same measurement limitation, it is reasonable to expect that application of UWI to bulk waves can yield information relevant to NDE.

This chapter aims to describe UWI in the context of NDE and demonstrate its application via several case studies; it is an expanded version of the conference proceedings by Michaels (2017). Advancement of UWI methods is rapid with many researchers worldwide actively engaged in developing methods, so the material presented here should not be considered as an exhaustive treatment but rather as an introduction. This chapter is organized as follows. Practical information on the acquisition and visualization of wavefield data is first discussed. Then, several analysis methods are described and illustrated using experimental wavefield data. Next, three case studies are presented that relate to detection and characterization of damage, two with guided waves and one with bulk waves. A brief review of other recent work is then provided to augment what is presented here. Finally, concluding remarks are made summarizing the status of wavefield imaging as an NDE method.

---

## Acquiring and Visualizing Ultrasonic Wavefield Data

The most common setup for acquiring wavefield data is to affix a piezoelectric transducer to the specimen of interest and record the resulting wave motion with a laser Doppler vibrometer (LDV). The LDV can either be attached to an XY scanner or it can be part of a scanning mirror system that directs the laser beam by tilting a mirror. A scanner-based LDV system is usually assembled by purchasing and integrating components, which include the scanner, a single-point laser vibrometer, a digitizer, and equipment such as a pulser-receiver or function generator to excite the wave source. A computer and appropriate software are required to complete the system. A mirror-based system is usually procured as a complete system including software that does not require integration or customization. Regardless of the specific LDV system, the resulting data consists of a set of signals that are proportional to the displacement or velocity of the wave motion, which for the most commonly used single-head LDV is out-of-plane motion. Three-axis LDVs are also available that measure all three displacement components, but the additional complexity and cost of these systems make them much less practical for use outside of the laboratory.

For either the scanner-based or mirror-based single-head LDV configuration, it can be challenging to obtain a good signal-to-noise ratio (SNR), and careful surface preparation is usually required. For lower frequencies (i.e., under 1 MHz), many laboratories have successfully used retro-reflective automotive tape, which is applied to the surface on which the laser vibrometer acquires data. This type of tape is particularly suited for a mirror-based LDV system because the laser beam is not perfectly normal to the surface over the entire scanning area. For a scanner-mounted LDV, the normality of the laser beam can be better controlled and other surface preparations that provide a strong specular reflection can be used. These can include surface polishing (for metallic specimens), reflective paint, and other types of reflective tape. Signal averaging is also typically applied to further improve the SNR, but there is a trade-off with acquisition time. As expected, for higher frequencies the presence of any additive surface preparation can disturb the wave propagation and hence affect the recorded signals.

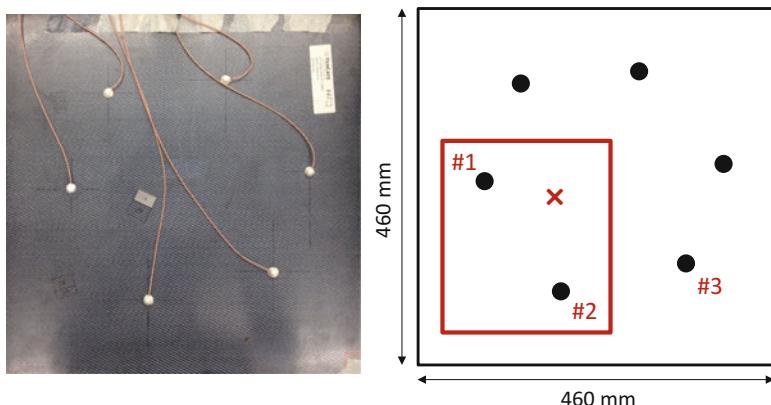
Another relatively common means of obtaining wavefield data is to use an air-coupled transducer as a receiver instead of the LDV, which has the advantages of lower cost and faster scan speeds, but presents several disadvantages, mainly related to data quality because of the large impedance mismatch between air and the specimen under test. Also, an air-coupled system does not directly measure surface motion but records airborne waves that have leaked from the specimen and propagated to the transducer. Completely non-contact systems have also been developed in which waves are generated by a fixed laser or air-coupled transducer and measured by a scanned LDV. An alternative approach is to generate waves with a scanned laser and receive with either a fixed piezoelectric transducer or a fixed LDV; the waves appear to originate at the fixed receiver location due to reciprocity of transmission and reception for linear elastodynamic systems. The laser source can also be scanned and the receiver fixed. These laser-based noncontact systems are described and compared by An et al. (2013). Lee et al. (2014) evaluated combined laser and air-coupled noncontact systems in which a scanned laser source generates waves that are measured with either a piezoelectric or capacitive air-coupled transducer. Although the use of a laser to generate ultrasonic waves has several advantages, it presents the disadvantages of safety issues, possible surface damage, and additional system complexity.

Ultrasonic wavefield data are by definition acquired in both time (one dimension) and space (one or two dimension). When acquired in one spatial dimension, the wavefield data can be represented as  $w(t,r)$ , where  $t$  is time and  $r$  is the single spatial dimension. A series of time signals are typically acquired along a line, although any curve could be the acquisition path, and these 2-D wavefields are referred to as either line scans or B-scans. An ultrasonic wavefield acquired in two spatial dimensions can be represented as  $w(t,x,y)$ , where  $x$  and  $y$  are the two spatial dimensions on the surface of the specimen, and these 3-D wavefields are referred to as area scans. The signals comprising the wavefield are typically acquired on a rectilinear grid with equal spatial sampling increments in both spatial dimensions. Line scans can be readily extracted from area scans.

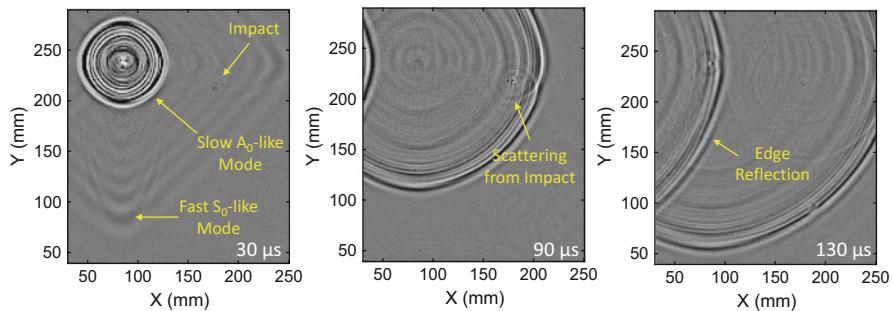
The simplest analysis method is wavefield visualization in two dimensions in the form of images. For line scans, the two dimensions are time and space, and the images show the trajectories of the measured waves. For area scans, data can be visualized by extracting line scans along any desired direction, but it is difficult to visualize the entire wavefield in this manner. The most effective visualization method for area scans is to view the wavefield as a movie. Although data are acquired by repeatedly exciting the fixed source and measuring the response at different spatial locations, the entire wavefield is best visualized by showing a series of 2-D images as time progresses, enabling the propagating wavefield to be viewed as a movie. Each frame is an  $x$ - $y$  image at a fixed time, which is also called a “snapshot.”

Figure 1 shows both a photograph and a drawing of an impact-damaged composite panel specimen from which laser vibrometer data were acquired. The panel measured  $460 \times 460 \times 2.5$  mm and was nominally homogeneous with an orthotropic layup of  $[0/90]_4$ . The photograph of Fig. 1a shows the transducer side of the panel on which six lead zirconate titanate (PZT) disks were attached with epoxy and covered with a bubble-filled epoxy mixture for protection. These radially polarized disks with a center frequency of 300 kHz were the sources of waves and were each 7 mm in diameter and 0.5 mm in thickness. Retro-reflective tape was applied to the other side of the panel over the area shown by the box in Fig. 1b, and laser vibrometer data were recorded over this  $219 \times 249$  mm region at an increment of 1.5 mm. Signals were digitized at 20 MHz and subsequently down-sampled to 5 MHz during post-processing. The total scan time was about 6 h, which works out to a little less than 1 s per point, and included extensive time averaging to increase the SNR.

Typical wavefield snapshots are shown in Fig. 2 for the case of transducer #1 as the source and with a 50–500 kHz chirp excitation. Data were processed prior to generating the plots to obtain the equivalent response to an impulsive excitation



**Fig. 1** Composite panel specimen (a) photograph and (b) diagram. The box indicates the wavefield scan area and the “ $\times$ ” designates the impact location



**Fig. 2** Wavefield snapshots from the composite panel specimen at times of (a) 30  $\mu$ s, (b) 90  $\mu$ s, and (c) 130  $\mu$ s

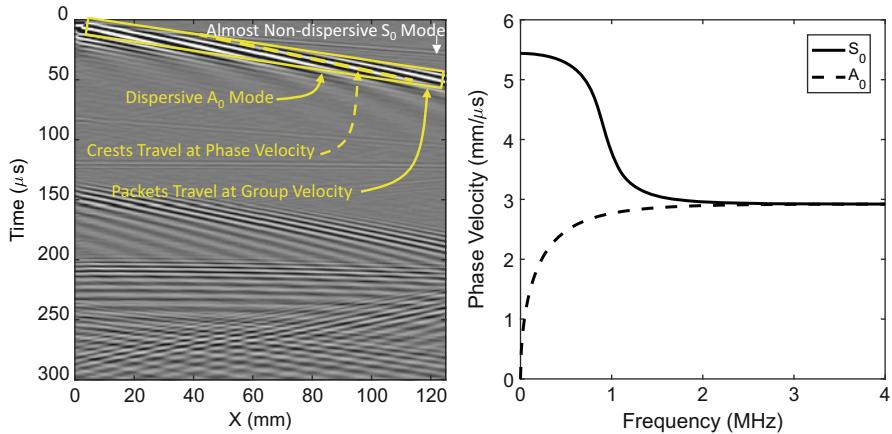
within the chirp bandwidth (Michaels et al. 2013). In the 30  $\mu$ s snapshot, it can be seen that the weak faster mode is very anisotropic with almost square wavefronts. In the 90  $\mu$ s snapshot, the slower mode has just passed the impact site and both scattering and energy trapping are evident. In the 130  $\mu$ s snapshot, a reflection of the slower mode from the left edge can be seen. Although scattering from the impact site is clearly visible, the wavefield is dominated by both the incident waves and edge reflections. Visualization of UWI data as either a series of snapshots or a movie provides a very useful and intuitive means of understanding how waves propagate in a specimen, but such visualization does not provide a quantitative assessment of damage.

## Analyzing Ultrasonic Wavefield Data

A variety of analysis techniques can be applied to UWI data with the goal of extracting quantitative NDE results. The concepts of phase and group velocity are first reviewed as applied to both guided and bulk UWI data. Then two energy imaging methods are defined and demonstrated using the composite panel data. Finally, Fourier domain (frequency-wavenumber) methods are described and illustrated, also with the composite panel data.

### Phase and Group Velocities

When analyzing guided wavefields, one must be cognizant of the difference between group velocity and phase velocity, which is caused by geometric dispersion (Rose 1999). Figure 3a shows LDV data acquired from a 1-D scan (in space) of guided waves propagating in a 2.54 mm thick aluminum plate. In this B-scan (time-distance) presentation, the time axis does not correspond to depth within the part, as is the case for a conventional B-scan, but to the propagation time from the source to the spatial location. The two fundamental Lamb wave modes are present, the faster  $S_0$  mode and the slower  $A_0$  mode, and their theoretical phase velocity



**Fig. 3** (a) Wavefield data recorded from a 2.54 mm thick aluminum plate, and (b) the theoretical dispersion curves for the two fundamental guided wave modes

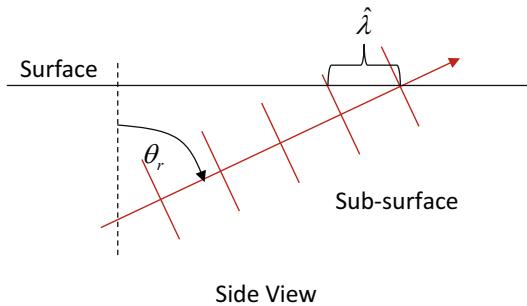
dispersion curves are shown in Fig. 3b. The  $S_0$  mode appears much weaker than the  $A_0$  mode because its out-of-plane displacement on the measurement surface is quite small compared to its in-plane displacement. The  $S_0$  mode is largely nondispersive with the group velocity of the wave packet traveling at essentially the same speed as the peaks and valleys, which travel at the phase velocity. The slower  $A_0$  mode is strongly dispersive, and it can be clearly seen that the wave packet is traveling faster than the peaks and valleys; i.e., the group velocity  $c_g$  is faster than the phase velocity  $c_p$ .

The situation is different when acquiring bulk wave signals, which propagate in the bulk material and are not guided by the specimen surfaces. Geometric dispersion does not apply to bulk waves, and the direction of propagation is typically either oblique or normal to the surface rather than parallel to the surface as is the case for guided waves. For example, if wavefield data are recorded from normally incident longitudinal waves, the entire wavefront hits the measurement surface at the same time, resulting in an apparent phase velocity of infinity. In general, for a refracted angle of  $\theta_r$ , as illustrated in Fig. 4, the apparent wavelength on the surface is elongated and the apparent phase velocity on the surface is increased by a factor of  $\sin(\theta_r)$ ; that is,

$$c_p = \frac{c}{\sin(\theta_r)}, \quad (1)$$

where  $c$  is the shear or longitudinal bulk wave speed and  $c_p$  is the apparent phase velocity on the specimen surface. The apparent group velocity on the surface is generally not of interest for bulk wavefield data since data are not measured along the direction of propagation and the apparent group velocity is generally a function of measurement location.

**Fig. 4** Apparent wavelength elongation of bulk waves caused by oblique propagation



## Energy Imaging

The entire wavefield is readily visualized as a movie, and although the viewer can obtain qualitative insights from the movie, it is desirable to obtain quantitative information similar to that which can be found in a conventional ultrasonic C-scan. Individual snapshots are such an image, but they do not capture information from the entire wavefield. A straightforward alternative is to generate an energy image  $E(x,y)$ ,

$$E(x,y) = \sum_{t=t_1}^{t_2} w^2(t,x,y), \quad (2)$$

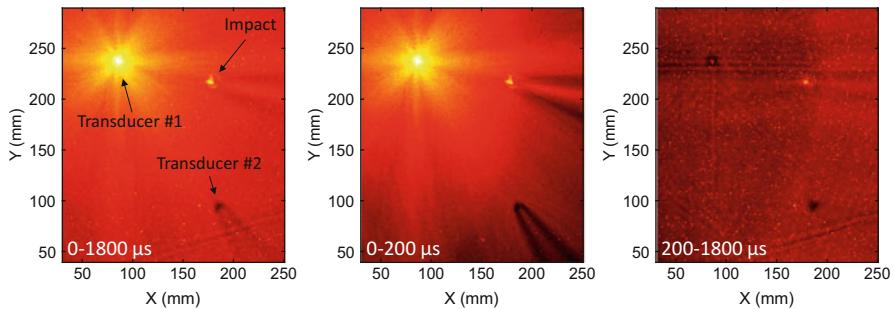
where  $t_1$  and  $t_2$  are the start and stop times for a time window in which energy is accumulated.

If the data are broadband, it may be useful to first transform the time-windowed data into the frequency domain and accumulate energy in a defined frequency range. If  $W(\omega,x,y)$  is the 1-D Fourier transform of the time-windowed wavefield, then the energy image corresponding to any desired frequency band can be directly computed in the frequency domain,

$$E(x,y) = \sum_{\omega=\omega_1}^{\omega_2} W^2(\omega,x,y), \quad (3)$$

where  $\omega_1$  and  $\omega_2$  are the start and stop radian frequencies ( $\omega = 2\pi f$ ) for the frequency range over which energy is accumulated.

Figure 5 shows three energy images of the composite panel shown on the same 30 dB scale that were computed using three different time windows. The image computed using the full time window of 0 to 1800  $\mu$ s clearly shows the impact damage as well as the shadowing effect of transducer #2, although the high-energy region around the source transducer dominates. The image computed using the narrow time window of 0 to 200  $\mu$ s is similar although the shadowing effects of both the impact damage and transducer #2 are more evident. This time window was selected to ensure that the primary incident wave from the source transducer reached the lower right corner of the scan area so that the entire area of interest is interrogated but edge reflections are minimized. The third image was computed using the 200 to 1800  $\mu$ s time window,



**Fig. 5** Energy images of the composite panel specimen computed for three different time windows

which does not include the primary incident wave but does include multiple edge reflections. Each location on the panel is thus interrogated by multiple waves that are incident from different directions. Both transducers appear as anomalies of lower energy with no obvious shadowing, and the impact appears as a higher energy anomalous region because of energy trapping within the delamination. The noisy background is typical of laser vibrometer data and is caused by laser speckle that is more evident when signal levels are low. In general, energy imaging tends to highlight inhomogeneous regions where an otherwise smoothly varying energy field is disturbed, and is thus most effective for identifying anomalies in otherwise homogeneous parts.

## Direct Arrival Imaging

A wavefield is typically recorded over a long enough time window for the waves to propagate over the entire region of interest. At all or some spatial locations, received signals typically include not only the direct arrival from the source but reflections and mode-converted signals from boundaries and other geometrical features of the specimen. An energy image that is constructed over a large time window includes all of these contributions, which can be useful since defects may be interrogated multiple times by reflected and scattered waves. A smaller time window focuses on the waves arriving directly from the source, but one window may not be optimal for all spatial locations.

An alternative approach is to create an energy image from only the direct arrival of the source waveform, which is the first wave to interact with a specific spatial location. If the source is located at  $(x_s, y_s)$  and the image pixel is located at  $(x, y)$ , the arrival time  $t_{\text{arr}}$  can be calculated as

$$t_{\text{arr}}(x, y) = \frac{\sqrt{(x - x_s)^2 + (y - y_s)^2}}{c_g} + t_{\text{cal}}, \quad (4)$$

where  $c_g$  is the group velocity and  $t_{\text{cal}}$  is a calibration offset time. An energy image can be calculated using Eq. (2) but with  $t_1$  and  $t_2$  being a function of  $x$  and  $y$  such that the time window encompasses the direct arrival:

$$t_1(x,y) = t_{\text{arr}}(x,y) - \Delta t/2 \quad \text{and} \quad t_2(x,y) = t_{\text{arr}}(x,y) + \Delta t/2. \quad (5)$$

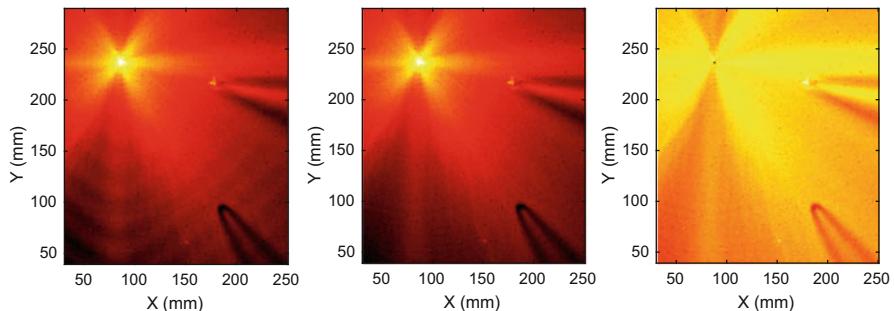
The total window width of  $\Delta t$  is set so that the time window for energy accumulation includes the directly arriving wave packet from the source. Here  $c_g$  is assumed to be constant, which may not be the case. If the specimen is anisotropic, then  $c_g$  would be a function of propagation direction. If the specimen is inhomogeneous, then  $c_g$  may be a function of position and it could be challenging to calculate the arrival time for each pixel. Direct arrival imaging highlights how the first arrival is perturbed by not only any inhomogeneity at the pixel location but also by the path of the wave from the source to that location.

Geometric spreading losses as well as material attenuation cause direct arrival amplitudes to decrease as a function of distance from the source. An amplitude correction can be applied to compensate for these losses,

$$E(x,y) = \sum_{t=t_1}^{t_2} w^2(t,x,y) \frac{d}{d_{\text{ref}}} e^{\alpha d}, \quad (6)$$

where  $d$  is the distance from the source to the point  $(x,y)$ ,  $d_{\text{ref}}$  is a reference distance,  $\alpha$  is the attenuation, and the times are calculated as per Eq. (5).

Figure 6 shows three direct arrival images constructed for the composite panel assuming a constant group velocity of 1.38 mm/ $\mu$ s and for a time window of 10  $\mu$ s. An additional calibration offset time of 2.2  $\mu$ s was applied to ensure best tracking of the first arrival. Figure 6a was computed used the raw (non-envelope-detected) signals and no amplitude correction. Note the waviness in the lower part of the image, which is caused by signal peaks “rolling” through the time window due to dispersion (mismatched phase and group velocities). Figure 6b was computed from the envelope-detected signals, which removes the waviness but is otherwise very similar, and Fig. 6c applies a correction for geometric spreading but not material attenuation. All images are shown on a 30 dB scale normalized to the maximum for the image.



**Fig. 6** Direct arrival images for the composite panel specimen. (a) Raw signals, no amplitude correction, (b) envelope-detected signals, no amplitude correction, and (c) envelope-detected signals, correction for geometric spreading

The direct arrival image of Fig. 6b is very similar to the energy image of Fig. 5b constructed using the narrow time window of 0 to 200  $\mu\text{s}$ . This similarity is not surprising since the panel is nominally homogeneous except for the impact damage and affixed transducers. As can be seen in the snapshots of Fig. 2, there is little energy behind the direct arrival prior to the arrival of edge reflections. The direct arrival image of Fig. 6c includes a correction for geometric spreading, which evens out the background energy level to some extent. All three images of Fig. 6 show that the angular pattern of the incident wavefield is not uniform with lower energy in the  $\pm 90^\circ$  directions and higher energy at  $0^\circ$  and  $180^\circ$ . These variations could be caused by transducer asymmetry, specimen anisotropy, or a combination of both.

## Frequency-Wavenumber Analysis

Wavefield data are converted from the time-space domain to the frequency-wavenumber domain using either the 2-D Fourier transform (for line scans) or the 3-D Fourier transform (for area scans). Since data are acquired in both time and space, both temporal and spatial Nyquist criteria must be met to avoid aliasing. In the time domain, the sampling frequency must be more than twice that of the highest frequency present in the signals, which is the same as stating that the temporal sampling increment must be less than half of the shortest period. In reality, data are typically oversampled, at least in the laboratory, at five to ten times the Nyquist rate to provide a smooth representation of the signals. Since such digitizers are readily available for the ultrasonic frequency ranges typical for ultrasonic NDE and memory is relatively inexpensive, such oversampling is usually not a burden. Such is not the case, however, for spatial sampling. To meet the spatial Nyquist criterion, the spatial sampling increment must be less than half of the smallest wavelength of propagating waves. Since the acquisition time is proportional to the number of waveforms acquired, halving the spatial sampling increment quadruples the acquisition time for an area scan. If the spatial Nyquist criterion is not met, the data may still be useful but it will likely not be possible to apply frequency-wavenumber methods.

In two dimensions, the continuous Fourier transform is defined as

$$W(\omega, k) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(t, r) e^{-i(kr - \omega t)} dt dr, \quad (7)$$

where  $r$  is the single spatial dimension,  $\omega$  is the angular frequency, and  $k$  is the wavenumber. The angular frequency, wavenumber, and phase velocity are related by

$$c_p = \frac{\omega}{k}. \quad (8)$$

In three dimensions, the continuous Fourier transform is

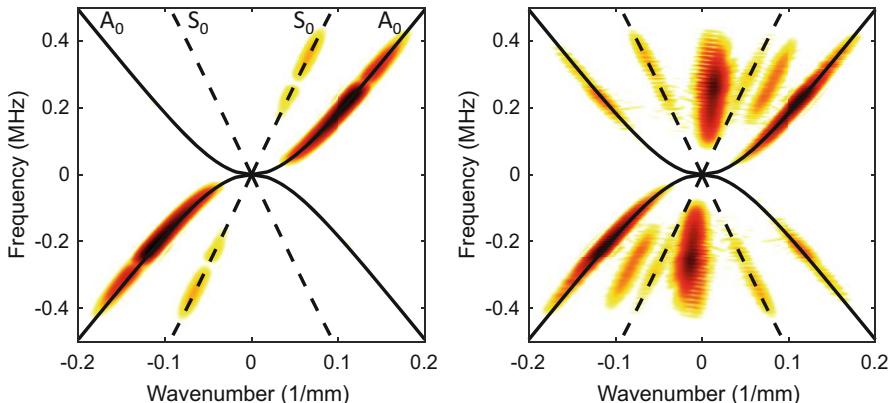
$$W(\omega, k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(t, x, y) e^{-i(k_x x + k_y y - \omega t)} dt dx dy, \quad (9)$$

and the wavenumber  $k = \sqrt{k_x^2 + k_y^2}$ . In practice, the discrete Fourier transform, rather than the continuous Fourier transform, is implemented using either the 2-D or 3-D fast Fourier transform (FFT) computational algorithm.

Figure 7 shows the 2-D frequency-wavenumber image of the guided wave signals shown in the B-scan of Fig. 3a for two different time windows. For both cases, a Hanning window is applied in both time and space to minimize spectral leakage. In Fig. 7a, a narrow time window of 0 to 80  $\mu\text{s}$  is used that contains only the first arrivals of the two modes. It can be seen in this figure that the trajectories of both modes closely agree with the theoretical dispersion curves, which are superimposed on the image. A larger window of 0 to 300  $\mu\text{s}$  is considered in Fig. 7b, and although the dispersion curves are still evident, the presence of the edge reflections adds considerable complexity to the frequency-wavenumber image. In particular, the apparent phase velocities of edge reflections are higher than the actual phase velocities when the measurement direction of the line scan is not coincident with the propagation direction. This issue, which only occurs for the 2-D Fourier transform, is illustrated in Fig. 8. The measured phase velocity  $\hat{c}_p$  is increased from that of the propagating waves by the cosine of the angle between the two directions,  $\alpha$ :

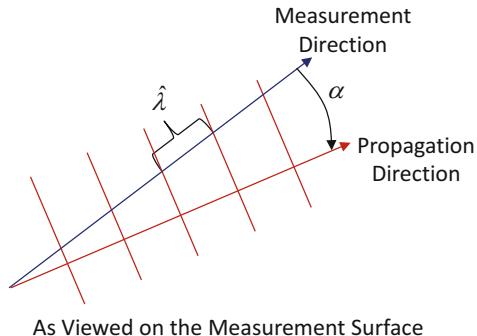
$$\hat{c}_p = \frac{c_p}{\cos(\alpha)}. \quad (10)$$

The 3-D frequency-wavenumber domain is similar to that in 2-D but since an area scan includes all directions of propagation, there cannot be a mismatch between



**Fig. 7** Frequency-wavenumber representation of wavefield line scan data recorded from a 2.54 mm thick aluminum plate. (a) 80  $\mu\text{s}$  window and (b) 300  $\mu\text{s}$  window

**Fig. 8** Apparent wavelength elongation caused by a mismatch in measurement and propagation directions



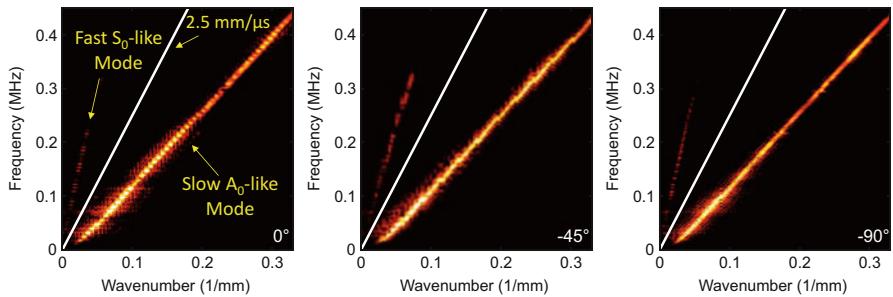
the measurement and propagation directions. Each discrete point in the 3-D frequency-wavenumber domain has corresponding values of  $\omega$ ,  $k_x$ , and  $k_y$ . Recalling that  $k = \sqrt{k_x^2 + k_y^2}$  and  $c_p = \omega/k$ , each location also has a well-defined phase velocity. In addition, the wavenumber vector  $\mathbf{k}$  points along the direction of propagation  $\hat{\mathbf{k}}$ :

$$\mathbf{k} = k_x \hat{\mathbf{e}}_x + k_y \hat{\mathbf{e}}_y \quad \text{and} \quad \hat{\mathbf{k}} = \frac{\mathbf{k}}{|\mathbf{k}|}. \quad (11)$$

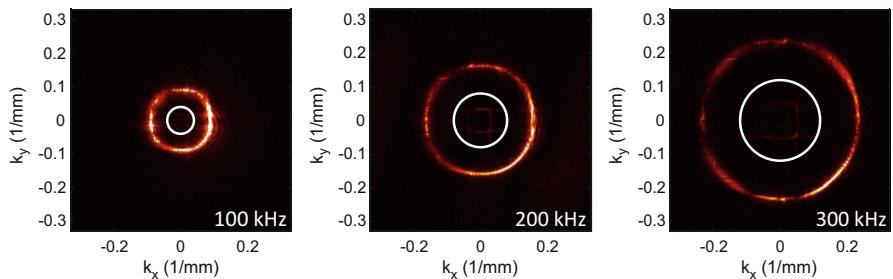
Thus, each point also has a well-defined propagation direction.

Figure 9 shows frequency-wavenumber slices for three different propagation angles after taking the 3-D FFT of the composite panel data. The entire wavefield was preprocessed to reduce the total time window to 700  $\mu\text{s}$  and smooth all edges in both time and space to avoid spectral leakage. These slices are similar to the one shown in Fig. 7a in that there is a dominant  $A_0$ -like slow mode and a much weaker  $S_0$ -like fast mode. The white line has a slope corresponding to a phase velocity of 2.5 mm/ $\mu\text{s}$  to serve as a reference; this line approximately separates the modes, although the actual mode shapes are not linear. The slope of the mode at a given frequency corresponds to the group velocity whereas the phase velocity is simply  $\omega/k$ . The slope of the fast mode at  $-45^\circ$  is less than at  $0^\circ$  and  $90^\circ$ , which is consistent with the fast mode wavefront shown in the snapshot of Fig. 2a. The slow mode is very slightly anisotropic but the group velocity of 1.38 mm/ $\mu\text{s}$  used for the direct arrival images of Fig. 6 matches well enough for accurate arrival time calculations within the region of interest for all propagation directions.

Figure 10 shows wavenumber-wavenumber ( $k_x - k_y$ ) slices for three different frequencies. The white circle in each figure again corresponds to a phase velocity of 2.5 mm/ $\mu\text{s}$ . The nearly isotropic slow mode appears as the larger diameter approximately circular shape whereas the weaker and strongly anisotropic fast mode corresponds to a square-like shape inside the 2.5 mm/ $\mu\text{s}$  reference circle; it is not visible on the 100 kHz plot and is very faint on the other two plots.



**Fig. 9** Frequency-wavenumber slices in the 3-D Fourier domain for the composite panel specimen and three propagation directions. (a)  $0^\circ$ , (b)  $-45^\circ$ , and (c)  $-90^\circ$

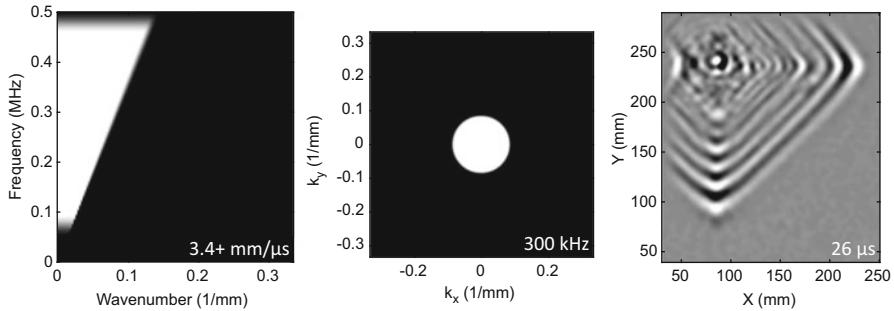


**Fig. 10** Wavenumber-wavenumber slices in the 3-D Fourier domain for the composite panel specimen and three frequencies. (a) 100 kHz, (b) 200 kHz, and (c) 300 kHz

## Frequency-Wavenumber Filtering

Direct visualization of wavefield data in the frequency-wavenumber domain is useful for understanding wave modes and overall propagation characteristics. But the real value of frequency-wavenumber analysis comes from filtering data in the Fourier domain and then transforming back to the time-space domain to analyze the filtered data (Ruzzene 2007; Michaels et al. 2011). Both phase velocity filtering and directional filtering are of interest to remove undesired phase velocities and propagation directions in the 3-D Fourier domain. An inverse FFT can then be applied to transform the filtered wavefield back to the time-space domain where it can be verified that the undesired wave components are no longer present.

As an example of phase velocity filtering, consider the various frequency-wavenumber domain slices shown in Figs. 9 and 10 for the composite panel. The weaker  $S_0$ -like mode is dominated by the much stronger  $A_0$ -like mode, but it can be enhanced by removing the stronger mode. Figure 11a, b shows two slices of a high-pass phase velocity filter that removes the strong mode and keeps the weak mode by setting a cutoff of  $3.4 \text{ mm}/\mu\text{s}$ . This filter additionally performs frequency-domain filtering by incorporating a bandpass filter from 50 to 500 kHz. The edges of the filter are smoothed to minimize spectral leakage. The snapshot of the filtered wavefield at



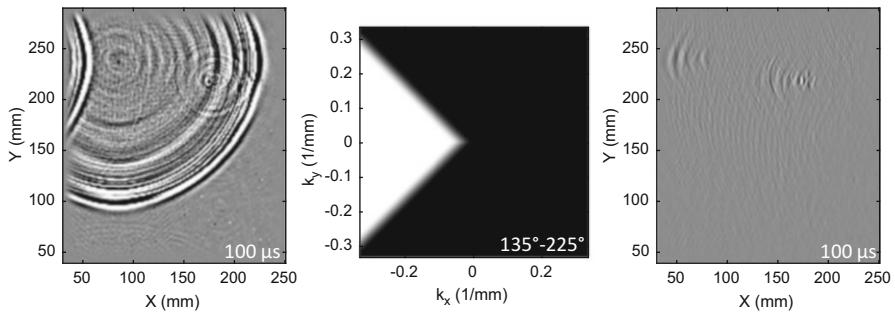
**Fig. 11** Illustration of phase velocity filtering of composite panel data from 3.4 mm/μs and faster. (a)  $\omega$ - $k$  filter slice, (b)  $k_x$ - $k_y$  filter slice, and (c) filtered snapshot at 26 μs

26 μs is shown in Fig. 11c where the anisotropic nature of this fast mode is much more evident than in the unfiltered snapshot of Fig. 2a. Although a linear phase velocity boundary was used here, the mode filters can also track the theoretical mode shapes as was done by Michaels et al. (2011) and Tian and Yu (2014).

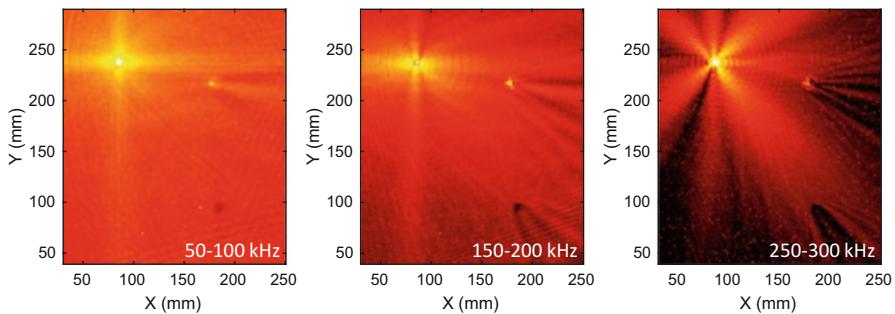
Figure 12 illustrates directional filtering by removing all waves propagating between  $-135^\circ$  and  $+135^\circ$ , keeping only those leftward propagating waves between  $135^\circ$  and  $225^\circ$ . Figure 12a shows a snapshot of the original wavefield at 100 μs with the gray scale set to enhance visualization of waves scattered from the impact damage. The details of these scattered waves are largely obscured by the incident waves. After application of the directional filter shown in Fig. 12b, the snapshot of Fig. 12c clearly shows the backscattered waves from the impact damage as well as leftward propagating waves from the source transducer.

## Guided Waves in a Composite Panel

The first case study continues the analysis of the wavefield data recorded from the composite panel as described in section “[Acquiring and Visualizing Ultrasonic Wavefield Data](#).” Results shown in section “[Analyzing Ultrasonic Wavefield Data](#)” indicate that the impact damage causes significant energy trapping whereas the effect of an attached transducer is to absorb energy. Figure 13 shows additional energy images computed as per Eq. (3) over the full time window of 1800 μs and for three different frequency ranges; all three images are shown on the same 30 dB scale. Figure 13a shows the image for a low-frequency band from 50 to 100 kHz. Although the image is dominated by the source transducer, both energy trapping and shadowing of the incident waves by the impact site are evident. The mid-frequency band image, from 150 to 200 kHz and shown in Fig. 13b, is also dominated by the source but displays a larger region of energy trapping and shows more of a disturbance of the wavefield from transducer #2; however, there is an overall higher background noise level. The highest band image shown, from 250 to 300 kHz and shown in Fig. 13c, continues the trend, but it is interesting to note that



**Fig. 12** Illustration of directional filtering of composite panel data from  $135^\circ$  to  $225^\circ$ . (a) Unfiltered snapshot at  $100\ \mu\text{s}$ , (b) directional filter, and (c) filtered snapshot at  $100\ \mu\text{s}$

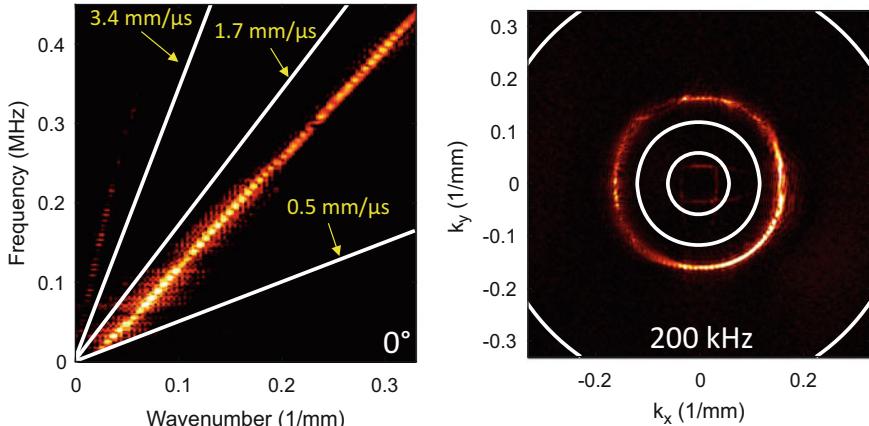


**Fig. 13** Narrowband wavefield energy images from the composite panel. (a) 50–100 kHz, (b) 150–200 kHz, and (c) 250–300 kHz

the energy pattern around the source transducer has changed quite a bit. The difference in patterns suggests that the layup of the plate or possibly the transducer itself has caused a frequency-dependent directional pattern of the interrogating wavefront. Although these frequency-domain energy images are useful, they are not comparable to conventional C-scans in terms of their ability to characterize damage.

Frequency-wavenumber analysis methods provide another set of tools to identify damaged regions. Figures 9 and 10 in section “[Frequency-Wavenumber Analysis](#)” illustrate the process of applying the 3-D Fourier transform to the wavefield (after time windowing and edge smoothing) to map it from the time-space domain to the frequency-wavenumber domain. In the frequency-wavenumber domain, the slow mode appears ring-like and the fast mode squarish at each frequency with the slower mode corresponding to larger diameter rings and the faster mode lying within the slower mode ring. Most of the energy of the wavefield is contained in these modes, which are dominant in the undamaged plate and include incident waves, edge reflections, and scattered waves.

A strategy for identifying anomalous regions, which might be damage, is to look at the distribution of energy that is not associated with these modes. Figure 14a shows a



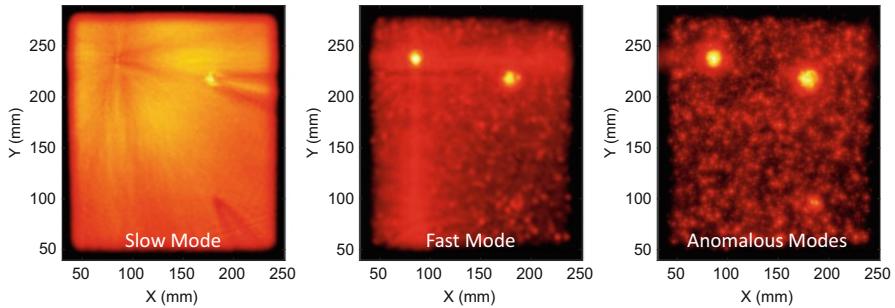
**Fig. 14** 2-D frequency-wavenumber slices showing phase velocity filtering boundaries. (a) Frequency-wavenumber slice at  $0^\circ$ , and (b) wavenumber-wavenumber slice at 200 kHz

frequency-wavenumber slice with the three white lines corresponding to phase velocities of 0.5, 1.7 and 3.4 mm/ $\mu$ s in order of increasing slope. The wedge-shaped region between 0.5 and 1.7 mm/ $\mu$ s bounds the dominant slow mode and the one between 3.4 mm/ $\mu$ s and infinity (vertical axis) bounds the weaker fast mode. The wedge between 1.7 and 3.4 mm/ $\mu$ s does not contain energy from either of the dominant modes; in fact, it appears to contain no energy at all, but that is not the case. Figure 14b shows a wavenumber-wavenumber slice at 200 kHz with the three circles corresponding to phase velocities of 0.5, 1.7, and 3.4 mm/ $\mu$ s in order of decreasing diameter.

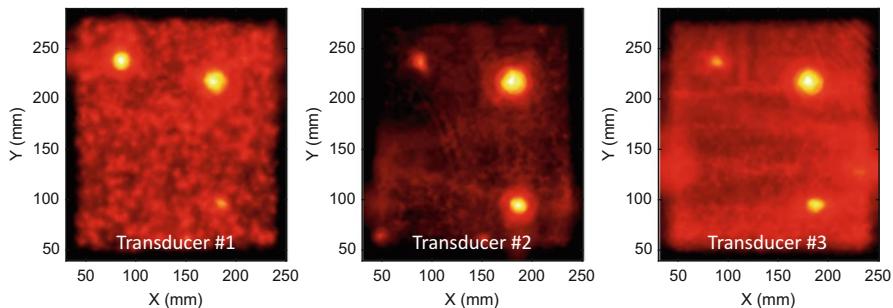
As previously described in section “[Frequency-Wavenumber Filtering](#),” phase velocity filtering can be applied to obtain filtered wavefields corresponding to three ranges of phase velocities: 0.5 to 1.7 mm/ $\mu$ s for the dominant slow mode, 3.4 to  $\infty$  mm/ $\mu$ s for the weaker fast mode, and 1.7 to 3.4 mm/ $\mu$ s for the so-called anomalous modes. Figure 11 shows the specific filter used for the fast mode, and similar filters were constructed for the other two ranges of phase velocity. Figure 15 shows energy images of these three wavefields, which confirm that the anomalous modes are indeed indicative of anomalies such as damage. The three indications on this image correspond to the site of impact damage and the two attached transducers, all of which are anomalies in the panel.

Wavefield data were also recorded over the same area using transducers #2 and #3 as sources, and “anomalous mode” images similar to that of Fig. 15c were generated from these two transducers. Figure 16 shows all three images after applying a  $3 \times 3$  median filter to reduce the speckle noise. Although the background noise level is different for the three images, they are remarkably similar and unambiguously detect the impact damage. The attached transducers also appear as anomalies, which is expected since they perturb the wavefield.

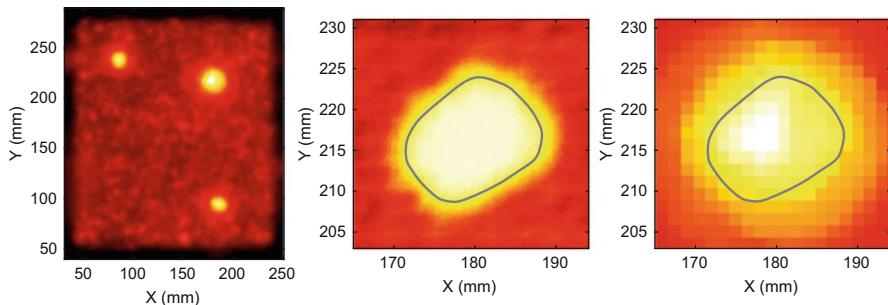
The three images were “fused” together by averaging to yield a composite image as shown in Fig. 17a, also shown on a 30 dB scale. To evaluate the efficacy of the



**Fig. 15** Wavefield energy images after frequency-wavenumber filtering. (a) Slow mode, (b) fast mode, and (c) anomalous modes. Each image is shown on a 30 dB scale relative to its peak value



**Fig. 16** Wavefield energy images after frequency-wavenumber filtering. (a) Slow mode, (b) fast mode, and (c) anomalous modes. Each image is shown on a 30 dB scale relative to its peak value



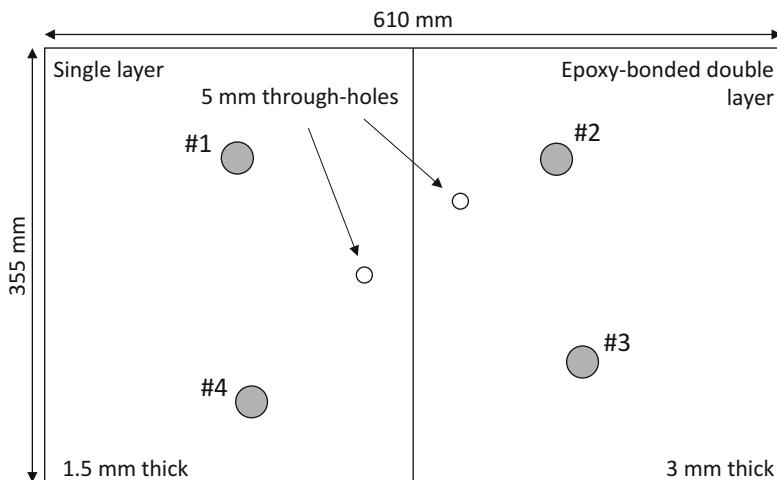
**Fig. 17** (a) Fused “anomalous mode” energy image, (b) double through-transmission C-scan, and (c) zoomed energy image. The curves in (b) and (c) are identical for comparison

wavefield imaging method, it can be compared to a conventional immersion C-scan, which is shown in Fig. 17b. This image was obtained with a 10 MHz, 12.4 mm diameter, 100 mm focal length immersion transducer using the double through-transmission method. In this method, also called the reflectoplate method, signals are

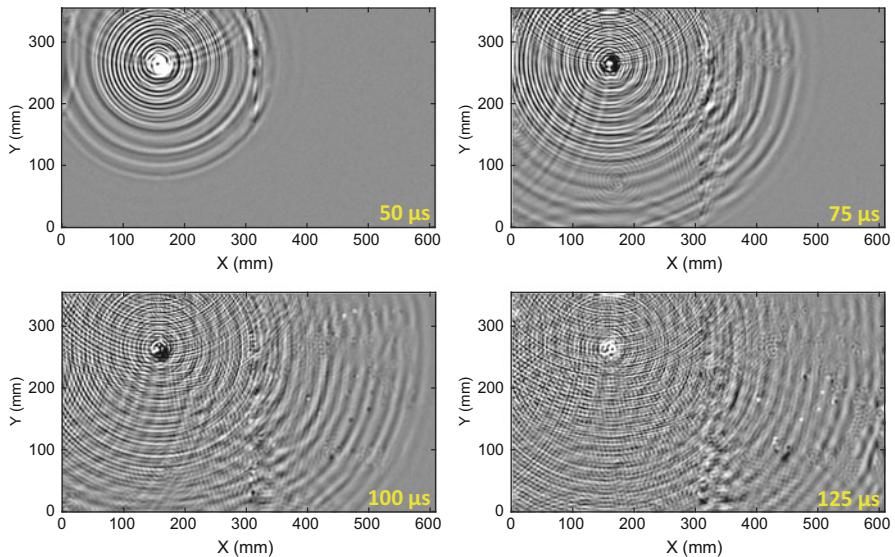
generated and received in pulse-echo mode, but the gate is set to report the peak amplitude from a far-side reflecting plate. Note that the color bar is inverted so that the bright region corresponds to low signal amplitudes caused by the impact blocking the longitudinal wave. Figure 17c is a zoomed version of Fig. 17a over the same area as the C-scan was performed. The curve tracing the impact damage, which was hand-drawn to match the indication in the C-scan, is shown in both Fig. 17b, c. It matches well with the wavefield anomalous mode energy image despite the fact that the resolution is much coarser, the color scales are not strictly comparable, and the spatial alignment between the two scans is not perfect.

## Guided Waves in a Bonded Plate

The second case study is that of guided waves propagating in a bonded aluminum plate; the experiment was originally described by Michaels and Michaels (2006). The specimen was constructed by bonding together two 1.5 mm thick aluminum plates with epoxy; the bond was intentionally fabricated with numerous air bubbles and an inconsistent thickness. As seen in the diagram of Fig. 18, the upper plate, which measured approximately 355 × 610 mm, was twice the size of the lower plate, which measured 355 × 305 mm. Four 12.5 mm diameter PZT disks were attached to the back side of the specimen, and transducer #1 was driven by an impulsive excitation to generate guided waves. The resulting wave motion was measured over the entire area at a spatial increment of 1.27 mm using a 400 kHz, 50 mm focal length, air-coupled transducer that was oriented normal to the upper specimen surface. Signals were sampled at 5 MHz over a 250  $\mu$ s time window beginning at the



**Fig. 18** Bonded plate specimen

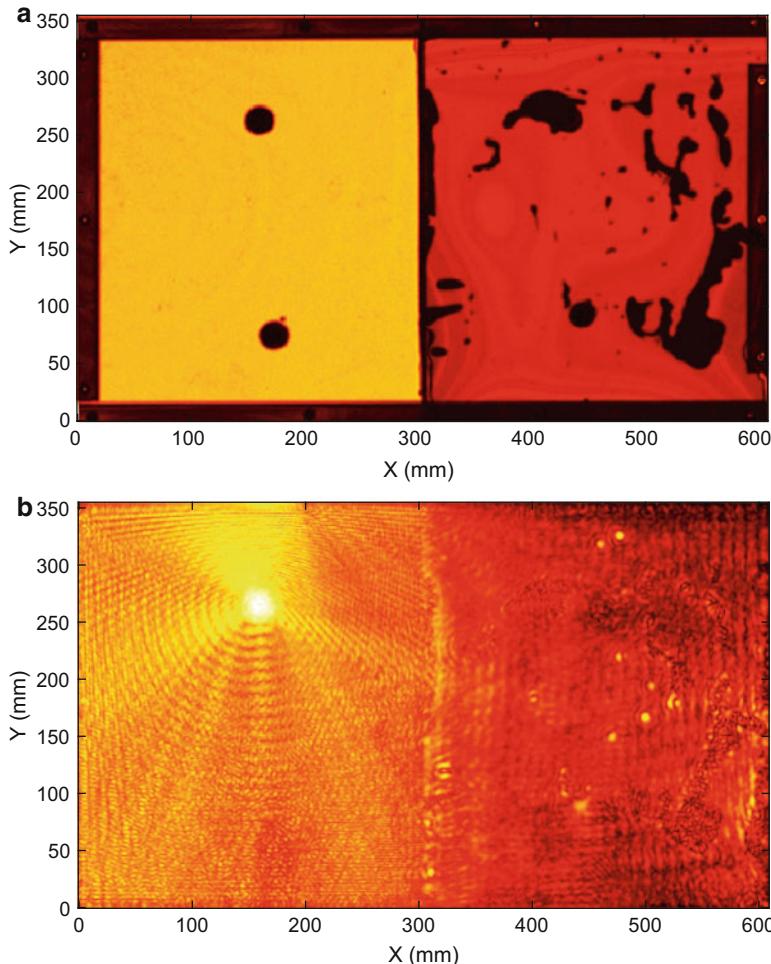


**Fig. 19** Wavefield snapshots from the bonded plate specimen

first arrival of the airborne wave at the receiver. The total scan time was about 20 min for the 134,400 signals, which is less than 10 ms per point.

Figure 19 shows snapshots of the measured wavefield at four different times. The zero-time reference is the time of arrival of the first wave at the receiver when the transducer is directly above the transmitter; that is, the propagation time through the air has been subtracted. In the snapshot at 50  $\mu$ s, the faster, longer wavelength  $S_0$  mode can be seen to lead the circular wavefronts followed by the slower, shorter wavelength  $A_0$  mode. As time progresses, the waves emanating from the source transducer appear to reverberate, which is actually due to the resonance of the narrowband air-coupled receiver. The waves can be seen to interact with the other attached transducers, the vertical boundary between the single plate on the left and the double layer on the right, and the various air bubbles in the epoxy layer. In addition, mode conversion from  $S_0$  to  $A_0$  at the attached transducers and the bond line can be readily observed. As time progresses, edge reflections cause the wavefield to become increasingly complicated, but also cause the epoxy layer air bubbles to become better delineated, primarily because of wave trapping within the bubbles.

Although the eye can readily discriminate most features of interest while playing the wavefield as a movie, it is more challenging to obtain a clear image of those features such as can be obtained by conventional ultrasonic imaging. Figure 20a shows a double-through-transmission C-scan of the entire specimen that was obtained with a 10 MHz, 12.4 mm diameter, 100 mm focal length immersion transducer. As expected, the bonding defects and the attached transducers can be clearly visualized as an almost complete loss of amplitude. Thickness variations are



**Fig. 20** (a) C-scan of bonded plate specimen, and (b) wavefield energy image

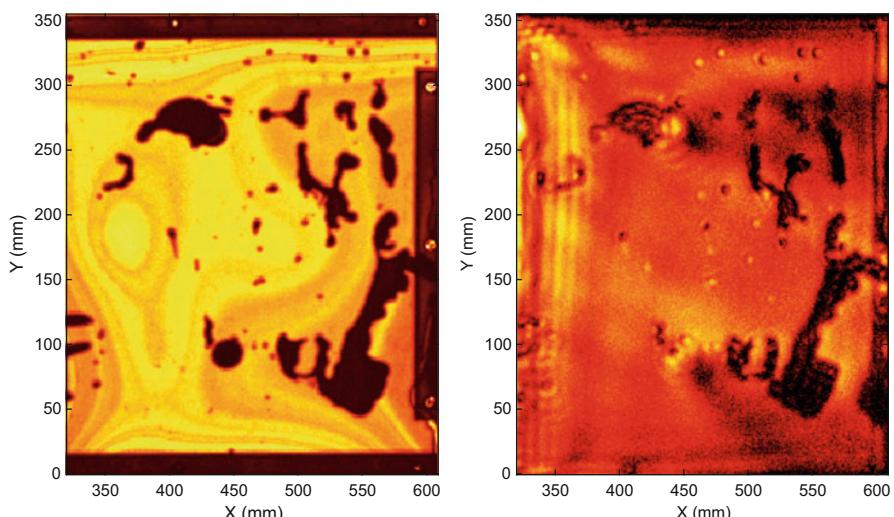
manifested as small changes in amplitude. The two through-holes were added after the C-scan was performed, so they do not appear in the image. Figure 20b shows a total energy image as obtained from the wavefield data for comparison; it is shown on a 30 dB scale. It might be expected that this image would highlight energy trapping in the bonding defects. However, although there are many indications in this image that correspond to the defects visible in the C-scan, most of them are not clearly delineated.

This wavefield does not lend itself to Fourier analysis for two main reasons. First, signals are saturated, particularly near the source transducer. Second, the reverberations of the receiver cause the wavefronts to appear to reverberate so that even if

modes and directions are separated, there is no resulting localized behavior in the time-space domain. Despite these complications, when viewing the wavefield movie, it can be seen that the leading edge of the faster  $S_0$  mode has the clearest interaction with the specimen features since there is little or no interference with either the slower mode or edge reflections. Although the specimen is not homogeneous, the wave speed of the fast mode is similar for the two halves of the bonded plate (single layer versus double layer). Thus, a direct arrival image can be readily calculated using Eq. (5) with a constant group velocity to calculate the arrival time of the fast mode at any position  $(x, y)$  on the specimen. If the group velocity changed significantly from one side of the plate to the other, the arrival time could still be calculated but with a more complicated expression.

Figure 21 shows a side-by-side comparison of the C-scan and the arrival time image for the right side of the specimen. Direct arrival imaging is not as effective for the left side because of signal saturation. The parameters for the arrival time image are  $c_g = 5.3 \text{ mm}/\mu\text{s}$ ,  $\Delta t = 10 \mu\text{s}$ , and  $t_{\text{cal}} = 20 \mu\text{s}$ , and it is shown on a 20 dB scale for best comparison to the C-scan, whose color scale has been adjusted to increase the dynamic range. The direct arrival image is remarkably similar to the C-scan with almost all indications appearing on both images. One difference is that some of the small indications, which are presumably air bubbles in the epoxy, are manifested as higher energy indications in the direct arrival image whereas they are lower in the C-scan. Correction for geometrical spreading has been applied, although some energy reduction is still evident, and the expected shadowing of the direct arrival by transducers #2 and #3 can be seen.

The direct arrival image is surprisingly good, particularly considering that the wavelength of the  $S_0$  mode is approximately 20 mm compared to about 0.6 mm for



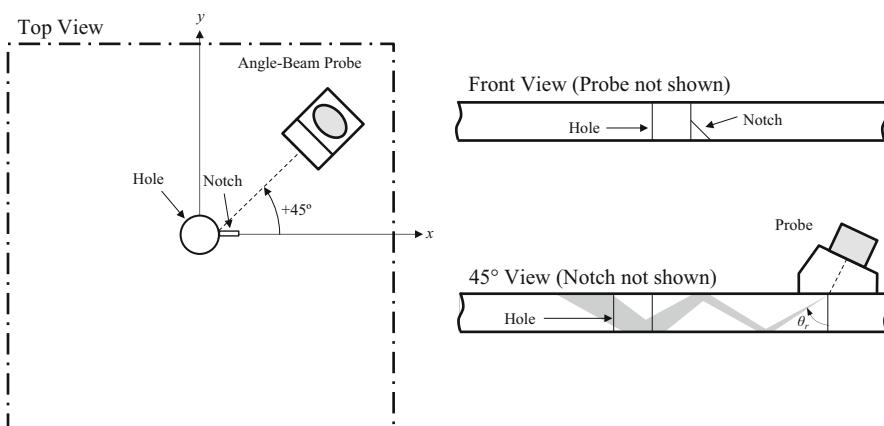
**Fig. 21** Images of the right side of the bonded plate. (a) C-scan and (b) direct arrival image

the 10 MHz C-scan. One of the strengths of air-coupled wavefield scanning is the very fast acquisition time, which was about two orders of magnitude faster than would have been the case for a laser vibrometer. The direct arrival image highlights the effectiveness of this analysis method for bonding defects even when the data quality is compromised by reverberations.

## Bulk Waves in an Aluminum Plate

The third case study applies wavefield imaging to investigate scattering of angle-beam shear waves from a notch emanating from a through-hole in an aluminum plate. Although wavefield imaging is particularly suited to the measurement of guided waves, it can also be applied to bulk waves, keeping in mind that the surface measurements are only a small portion of the total wavefield. The motivation of this study was to better understand scattering of angle-beam shear waves from back surface cracks with the long-term goal of improving inspection methods for such cracks.

Figure 22 illustrates the experimental configuration, in which a through-hole with a back-surface corner notch is interrogated with a conventional 5 MHz angle-beam shear wave probe with a nominal refracted angle of 56.8°. The thickness of the 6061 aluminum plate was 6.35 mm and the top surface had a mirror finish to improve the SNR. Wavefield data were acquired over a 30 × 30 mm square region centered about the 6.35 mm diameter through-hole using a laser Doppler vibrometer measuring out-of-plane displacement. The spatial increment in both  $x$  and  $y$  was 0.25 mm and the total scan time was about 3.5 h (about 1 s per point). The excitation was a 180  $\mu$ s, 200 Vpp linear chirp from 1 to 10 MHz, and received signals were post-processed via deconvolution to obtain the equivalent response to a 2-cycle, 5 MHz, Hann-



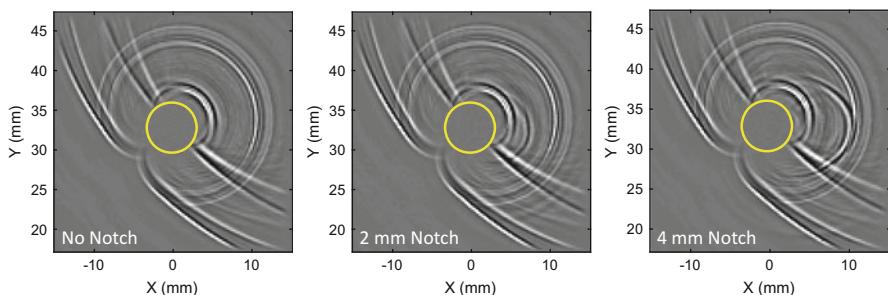
**Fig. 22** Sketch of the aluminum plate specimen showing the through-hole, back-surface corner notch, and angle-beam probe (not to scale)

windowed tone burst (Michaels et al. 2013). Data sets were recorded both before and after introduction of the hand-cut back-surface corner notch that was initially 1 mm in length and was subsequently enlarged to 4 mm.

Figure 23 shows three wavefield snapshots at a time of 16.88  $\mu\text{s}$  from the undamaged specimen and after introduction of the 2 mm and 4 mm notches. The through-hole boundary is shown on these and subsequent plots as an open circle. The waves are incident from the upper right as shown in Fig. 22. As can be seen in the plots, the wavefield has been spatially windowed to smooth the edges and also to remove the noisy signals obtained when scanning over the through-hole. A comparison of these snapshots clearly shows scattering from the notch, and an increase in scattering for the 4 mm notch as compared to the 2 mm notch. However, the wavefields are very complicated, consisting of incident waves, hole-scattered waves, and notch-scattered waves, and it is difficult to quantify the scattering from just the notch by simple spatial or temporal windowing. The hole-plus-notch is a compound scatterer and the waves scattered from the notch are particularly difficult to separate from those scattered from the hole.

The smallest wavelength in the data corresponds to the slowest wave, which is the Rayleigh wave whose nominal phase velocity is 2.9  $\text{mm}/\mu\text{s}$ . Even though the angle-beam probe is designed to generate only shear waves, there are Rayleigh waves present in the incident wavefield (Dawson et al. 2016). Given the spatial increment of 0.25 mm, the minimum wavelength to avoid spatial aliasing is 0.5 mm, which corresponds to a maximum Rayleigh wave frequency of 5.8 MHz. To minimize Rayleigh wave aliasing while keeping as much information as possible, a simple bandpass filter was applied to each waveform by multiplication in the frequency domain with a 33% Tukey window from 0.5 to 6.5 MHz.

The analysis approach taken here generally follows that described by Dawson et al. (2017), and consists of the following main steps: (1) wavefield baseline subtraction to isolate notch-scattered waves, (2) 3-D frequency-wavenumber filtering to isolate shear waves, (3) extraction of radial B-scans at different angles relative to an observation point, and (4) accumulation of energy in the 2-D frequency-wavenumber domain to generate a scattering pattern. The end result captures



**Fig. 23** Wavefield snapshots at 16.88  $\mu\text{s}$  for (a) no notch, (b) 2 mm corner notch, and (c) 4 mm corner notch

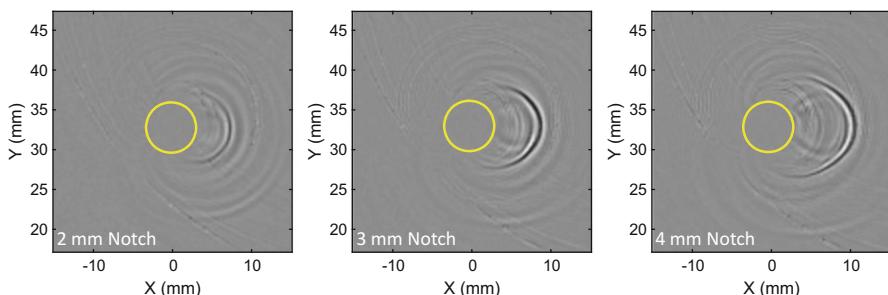
scattered energy from all of the shear wave skips, where one skip refers to a full “bounce” from the top surface of the plate to the bottom surface and back to the top.

The first step of wavefield baseline subtraction, as its name implies, is the subtraction of two wavefields to yield a residual wavefield,

$$w_r(t,x,y) = w_c(t,x,y) - w_b(t,x,y), \quad (12)$$

where  $w_c(x,y,t)$  is the current wavefield,  $w_b(x,y,t)$  is the baseline wavefield, and  $w_r(x,y,t)$  is the residual wavefield. Here the baseline wavefield is taken to be that recorded from the undamaged specimen (hole with no notch), and the current wavefield is after introduction and subsequent enlargement of the notch. Although wavefield baseline subtraction is straightforward in principle, in practice it is difficult to get acceptable results because it is virtually impossible to achieve perfect alignment between the current and baseline wavefields. Following the method developed by Dawson et al. (2016), each frame of the current wavefield is matched to a nearby frame of the baseline wavefield prior to subtraction where the baseline frame is spatially shifted to minimize the residual energy for each frame. Figure 24 shows residual wavefield snapshots at a time of 16.88  $\mu\text{s}$  for the 2 mm, 3 mm, and 4 mm notch sizes; these snapshots are shown on the same gray scale as those of Fig. 23. Although some feedthrough of the incident and hole-scattered waves is evident, baseline subtraction is largely successful in extracting the notch-scattered waves. The expected increase in notch scattering as the notch size increases can be clearly seen.

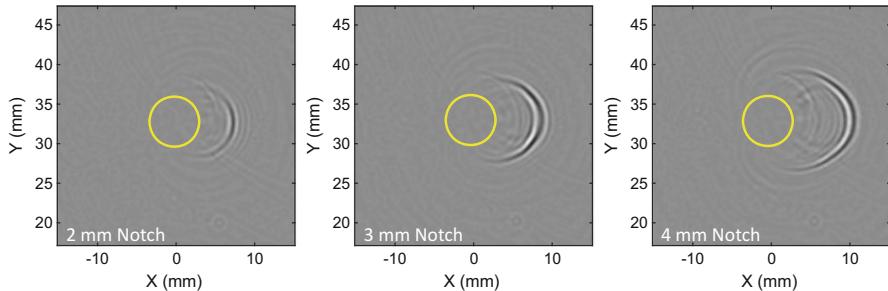
The second step is to apply phase velocity filtering in the 3-D frequency-wavenumber domain to extract shear waves. There are both incident and hole-scattered Rayleigh waves present in the wavefield as well as hole- and notch-scattered longitudinal waves (Dawson et al. 2016). Since scattered shear waves are of interest, it is desirable to remove the Rayleigh and longitudinal waves. Although the bulk shear wave speed is well-known, as described in section “Phase and Group Velocities” the apparent phase velocity on the surface is faster than the bulk wave speed by a factor of  $1/\sin(\theta_r)$ , where  $\theta_r$  is the refracted angle. Thus, the phase velocity limits for the phase velocity filter were set to the nominal shear and longitudinal wave speeds of 3.11 mm/ $\mu\text{s}$  and 6.32 mm/ $\mu\text{s}$ , respectively. Setting a higher upper limit would



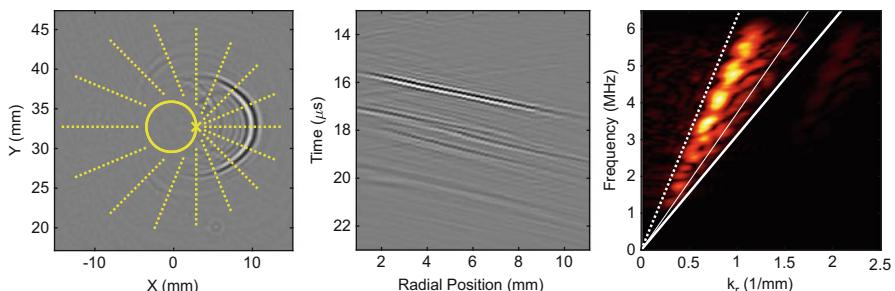
**Fig. 24** Residual wavefield snapshots at 16.88  $\mu\text{s}$  for three notch sizes. (a) 2 mm, (b) 3 mm, and (c) 4 mm

mean that some high-refracted-angle longitudinal waves could be included, and a smaller lower limit would include feedthrough of Rayleigh waves, whose nominal wave speed was 2.9 mm/ $\mu$ s. Figure 25 shows the three residual snapshots of Fig. 24 after phase velocity filtering, and it can be seen that they are significantly cleaner.

The third processing step is to extract radial B-scans from the residual wavefield relative to a specified reference point, which is illustrated in Fig. 26a. The residual wavefield snapshot in the background is that of Fig. 25c for the 4 mm notch. The reference point is designated by the “ $\times$ ” symbol, the hole is the solid circle, and the radial B-scan lines are the dotted lines, which are shown with a spacing of  $22.5^\circ$  for clarity. The reference point was chosen to be the hole-notch corner since the notch-scattered waves appear to approximately originate from this point. Since data are not measured inside the hole, the radial lines all begin outside of the hole and extend from 1 mm to 11 mm relative to the hole edge. Figure 26b shows the extracted radial B-scan for the 4 mm notch at an angle of  $0^\circ$ , and although there is some feedthrough of the incident waves, the outward propagating scattered waves dominate. Figure 26c shows the first quadrant of its frequency-wavenumber representation, which contains the forward (outward from the reference point) propagating waves.



**Fig. 25** Residual wavefield snapshots at 16.88  $\mu$ s for three notch sizes after phase velocity filtering. (a) 2 mm, (b) 3 mm, and (c) 4 mm



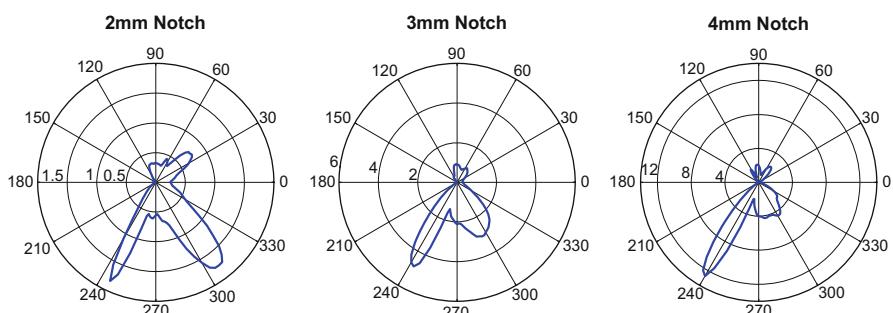
**Fig. 26** Extraction and frequency-wavenumber analysis of radial B-scans. (a) Radial B-scan lines originating from the hole-notch corner, (b) radial B-scan at  $0^\circ$ , and (c) frequency-wavenumber representation of  $0^\circ$  radial B-scan

The heavy, solid line corresponds to the shear phase velocity of 3.11 mm/μs and the heavy dotted line to the longitudinal phase velocity of 6.32 mm/μs; the thinner line at 3.72 mm/μs corresponds to the nominal refracted angle of 56.8°. It can be seen that there is very little energy outside of the shear range because of the phase velocity filtering step. It is interesting to note that most of the energy is in the phase velocity range between 3.72 and 6.32 mm/μs, which corresponds to refracted angles smaller than the nominal (i.e., approaching normal incidence).

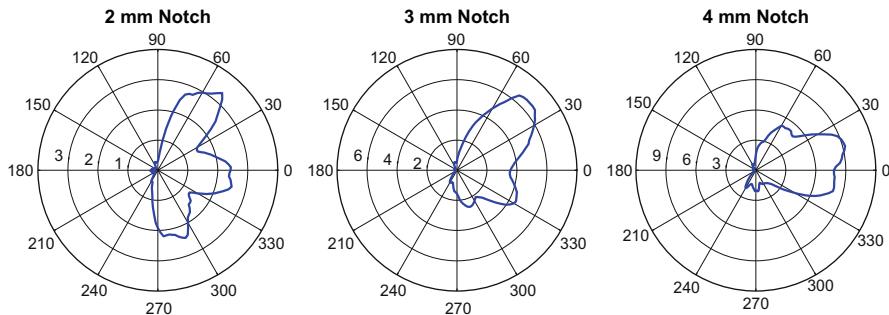
The final step is to accumulate energy in the frequency-wavenumber domain for each radial B-scan angle and for a specified phase velocity range. Different scattering patterns can be generated by specifying different ranges of phase velocities. Figure 27 shows 2, 3, and 4 mm notch scattering patterns computed for a phase velocity range of 3.62 to 3.82 mm/μs, which is a narrow range about the nominal phase velocity of 3.72 mm/μs (this range corresponds to refracted angles from 54.5° to 59.2°). These patterns all have narrow lobes with the largest lobe at about 240° (−120°), which corresponds to the notch shadowing the incident waves. All three patterns also have a strong but wider lobe at about 300° (−60°) and a third smaller lobe in the backscattered direction at close to +45°. Although the strength of all three lobes increases with notch size, their relative sizes change with the 2 mm notch having more nearly equal lobes than the 3 mm and 4 mm notches.

Figure 28 shows patterns for a broader but higher range of phase velocities, 4.4 to 6.3 mm/μs, which corresponds to a refracted angle range of 30–45°. These refracted angles are not present in the incident waves, at least not to a large degree, and are likely generated via diffraction from the notch edge or corner. The resulting scattering patterns are not as directional as those from the narrow range of phase velocities, which is consistent with the expected more omnidirectional scattering from a sharp discontinuity.

The methodology shown here for estimating scattering patterns is very general, and although applied here to bulk waves, is completely applicable to guided waves. In fact, its application to guided waves is more straightforward because there is no modal ambiguity as there is for shear and longitudinal bulk waves and it is easier to avoid spatial aliasing because of the generally larger wavelengths.



**Fig. 27** Shear scattering patterns for a narrow range of phase velocities about the nominal refracted angle. (a) 2 mm notch, (b) 3 mm notch, and (c) 4 mm notch



**Fig. 28** Shear scattering patterns for a range of phase velocities corresponding to refracted angles of 30–45°. (a) 2 mm notch, (b) 3 mm notch, and (c) 4 mm notch

## Additional Wavefield Methods

In addition to the wavefield methods shown here, there are many others that have been developed to address specific inspection problems. Several of these are reviewed in this section. In terms of visualization, a method similar to direct arrival imaging tracks a point of constant phase on the wavefield to image its variations over a region of interest. Unlike direct arrival imaging, the phase velocity must be used to track the wave crest instead of the group velocity. Michaels and Michaels (2007) called this method “wave crest amplitude analysis” and applied it to guided wave imaging of both simulated corrosion and notches in aluminum plates. Köhler et al. (2012) similarly tracked a crest on the wavefront of skimming longitudinal waves to map grain boundaries in an austenitic stainless steel weld.

Several analysis methods exploit the idea of anomalies in guided wavefields. Lee et al. (2012) proposed a time-space domain method whereby adjacent waves are time-aligned and subtracted to identify wavefield anomalies. Harb and Yuan (2016) applied frequency-wavenumber filtering to separate the wavefield into forward and backward propagating waves and applied a zero-lag cross-correlation method to identify anomalous backscattering caused by damage. Park et al. (2014) defined a standing wave filter based upon directional filtering and applied it to identify damage.

It is also possible to identify and characterize damage-induced anomalies in the wavenumber domain. Rogge and Leckey (2013) performed a local wavenumber analysis to not only detect impact damage in composites but to also estimate the depth of the impact. Flynn et al. (2013) similarly applied local wavenumber methods to estimate both wall thinning in metallic specimens and the depth of impact damage in composites. Kudela et al. (2015) performed wavenumber-wavenumber filtering in the spatial 2-D Fourier domain at each time slice to extract anomalous wavenumbers and then mapped each slice back to the spatial domain to identify damage. Yu et al. (2015) also noted the generation of anomalous wavenumbers by discontinuities and applied filtering using a short-space Fourier transform method to identify damage.

All of these methods illustrate the potential of wavefield imaging for a wide variety of inspection problems, but do not specifically address the typically very long scan times required for high-resolution measurements. Using air-coupled receivers is one possible solution which has the additional advantage of straightforward integration with conventional automated ultrasonic scanning systems but with the disadvantage of reduced data quality. A hardware-based approach is to use multi-point LDVs to measure multiple signals simultaneously (Kilpatrick and Markov 2010); such systems are in development but are not yet commercially available.

Another approach to address the scan time problem is to develop analysis methods that require fewer signals. Tian et al. (2016) proposed a “global-local” approach whereby the LDV system first acts as a receiver array to record a small number of signals that are used to construct a delay-and-sum image of a large area. Locations with anomalies are then targeted for high-resolution local scans, avoiding having to scan the entire specimen with a high resolution. Mesnil and Ruzzene (2016) took the approach of recording a small number of randomly distributed measurements and applying sparse reconstruction methods to locate both the actual source and secondary sources of waves that explain the measurements. These secondary sources are structural anomalies, or possible defects. Harley and Chia (2018) also considered a small number of randomly distributed measurements and used sparse reconstruction methods to estimate the damage-free wavefield. Residual signals at these random locations are then used to construct an image of damage.

One other promising approach to increase the speed of wavefield measurements is to apply a continuous excitation to generate a steady state wavefield at a constant frequency and measure the amplitude and phase with an LDV. Local wavenumber methods can then be applied to detect anomalies. Goodman et al. (2014) demonstrated this method to assess material distributions inside storage cylinders and achieved scan times of less than 10 s. Mesnil et al. (2016) similarly used continuous excitations to assess delaminations in composites, which reduced scan times from about 3 h to 10 min.

---

## Summary

There is no question that wavefield imaging is playing a critical role in understanding how ultrasonic waves propagate and interact with both damage and structural features, particularly for guided waves. The specific applications and methods shown here are representative of what can be achieved with UWI but are by no means a comprehensive treatment. Clearly UWI is moving from being a powerful laboratory tool to an effective NDE method.

Despite this demonstrated effectiveness, the practicality of wavefield imaging still must be addressed. There are at least three closely related obstacles that must be overcome for wavefield imaging to be implemented in the field: (1) scan time, (2) data quality, and (3) data analysis methods.

In terms of scan time, there are several approaches being considered to address this issue as discussed in section “[Additional Wavefield Methods](#).” Although the time-consuming LDV-based systems that record transient signals will likely serve as the

“gold standard” for some time, clearly there are other approaches, both hardware and software, that can be brought to bear to reduce scan time to more practical values.

Data quality, which is closely related to scan time, is also an issue. The quality of laser vibrometer data can always be improved by more signal averaging at the expense of scan time. Data quality is very dependent upon surface optical reflectivity, which is often improved in the laboratory by applying reflective tape or paint. However, disturbing the surface removes one of the primary advantages of wavefield imaging over conventional ultrasonic NDE methods, which require couplant. Using a continuous excitation significantly improves the data quality at the frequency(ies) being excited because it is possible to inject much more energy than a pulsed excitation, but this type of excitation limits analysis options.

Ultimately the success of wavefield imaging as an NDE method will depend upon the data analysis methods. Unlike conventional ultrasonic imaging, in which both transmitter and receiver are locally scanned to map discontinuities, in UWI the presence of either a defect or benign feature affects the global wavefield, not just measurements in the vicinity of the defect or feature. It is definitely challenging to extract local information despite global disturbances, and although research efforts thus far have been moderately successful in doing so, the efficacies of the various methods need to be quantified for a wide variety of structures and defects. Most, if not all, of the proposed analysis methods have in common the idea of recognizing anomalies in the wavefield that are caused by damage. This approach, which is similar to that of conventional UT inspections, works well when most of the specimen is homogeneous, but may fail for more complicated structures.

The main competition for UWI is conventional ultrasonic testing (UT). There needs to be a compelling reason to use UWI, and its noncontact nature is likely to be an important part of that reason. Another attractive feature of UWI is that only one-sided access is needed. Even if the quality of the inspection is not as good as can be achieved with conventional UT, if the overall inspection time is less, including any required disassembly and reassembly, UWI could be preferred. The future deployment of SHM systems may be the driving force to move wavefield imaging out of the lab and into the field.

**Acknowledgments** The first case study was supported by Advanced Systems & Technology, Inc., Irvine, CA, as part of a NASA SBIR Phase II Award, Contract No. NNX15CL26C, Dr. Cara Leckey, program manager. The third case study was sponsored by the Air Force Research Laboratory, Contract No. FA8650-10-D-5210, Dr. Eric Lindgren, program manager.

---

## References

- An YK, Park B, Sohn H (2013) Complete noncontact laser ultrasonic imaging for automated crack visualization in a plate. *Smart Mater Struct* 22:025022 (10pp)
- Dawson AJ, Michaels JE, Michaels TE (2016) Isolation of ultrasonic scattering by wavefield baseline subtraction. *Mech Syst Signal Process* 70–71:891–903

- Dawson AJ, Michaels JE, Kummer JW, Michaels TE (2017) Quantification of shear wave scattering from far-surface defects via ultrasonic wavefield measurements. *IEEE Trans Ultrason, Ferroelect Freq Control* 64(3):590–601
- Flynn EB, Chong AY, Jarmer GJ, Lee JR (2013) Structural imaging through local wavenumber estimation of guided waves. *NDT&E Int* 59:1–10
- Gannon A, Wheeler E, Brown K, Flynn E, Warren W (2015) A high-speed dual-stage ultrasonic guided wave system for localization and characterization of defects. *Conf Proc Soc Exp Mech Ser* 7:123–136
- Goodman D, Rowland K, Smith S, Miller K, Flynn E (2014) Non-destructive examination of multiphase material distribution in uranium hexafluoride cylinders using steady-state laser Doppler vibrometry. *Conf Proc Soc Exp Mech Ser* 5:81–88
- Hall G (1977) Ultrasonic wave visualization as a teaching aid in non-destructive testing. *Ultrasonics* 15(2):57–69
- Harb MS, Yuan FG (2016) Damage imaging using non-contact air-coupled transducer/laser Doppler vibrometer system. *Struct Health Monit* 15(2):193–203
- Harley JB, Chia CC (2018) Statistical partial wavefield imaging using Lamb wave signals. *Struct Health Monit* 17(4):919–935
- Kilpatrick JM, Markov VB (2010) Full-field laser vibrometer for instantaneous vibration measurement and non-destructive testing. *Key Eng Mater* 437:407–411
- Köhler B, Barth M, Krüger P, Schubert F (2012) Grain structure visualization with surface skimming ultrasonic waves detected by laser vibrometry. *Appl Phys Lett* 101:074101 (3pp)
- Kudela P, Radzieński M, Ostachowicz W (2015) Identification of cracks in thin-walled structures by means of wavenumber filtering. *Mech Syst Signal Process* 50–51:456–466
- Lee JR, Chia CC, Park CY, Jeong H (2012) Laser ultrasonic anomalous wave propagation imaging method with adjacent wave subtraction: algorithm. *Opt Laser Technol* 44:1507–1515
- Lee JR, Sunuwat N, Park CY (2014) Comparative analysis of laser ultrasonic propagation imaging system with capacitance and piezoelectric air-coupled transducers. *J Intell Mater Syst Struct* 25(5):551–562
- Mesnil O, Ruzzene M (2016) Sparse wavefield reconstruction and source detection using compressed sensing. *Ultrasonics* 67:94–104
- Mesnil O, Yan H, Ruzzene M, Paynabar K, Shi J (2016) Fast wavenumber measurement for accurate and automatic location and quantification of defect in composite. *Struct Health Monit* 15(2):223–234
- Michaels JE (2017) Ultrasonic wavefield imaging: research tool or emerging NDE method? In: Chimenti DE, Bond LJ (eds) Review of progress in quantitative nondestructive evaluation, vol 36. AIP conference proceedings 1806 (020001, 14pp)
- Michaels TE, Michaels JE (2006) Integrated monitoring and inspection with attached ultrasonic transducers. In: Kundu T (ed), Proceedings of the SPIE, vol 6177, 61770E (12pp)
- Michaels TE, Michaels JE (2007) Monitoring and characterizing corrosion in aluminum using Lamb waves and attached sensors. In: Kundu T (ed) Proceedings of the SPIE, vol 6532, 65321G (11pp)
- Michaels TE, Michaels JE, Ruzzene M (2011) Frequency-wavenumber domain analysis of guided wavefields. *Ultrasonics* 51:452–466
- Michaels JE, Lee SJ, Croxford AJ, Wilcox PD (2013) Chirp excitation of ultrasonic guided waves. *Ultrasonics* 53:265–270
- Nishizawa O, Satoh T, Lei X, Kuwahara Y (1997) Laboratory study of seismic wave propagation in inhomogeneous media using a laser Doppler vibrometer. *Bull Seismol Soc Am* 87(4):809–823
- Park B, An YK, Sohn H (2014) Visualization of hidden delamination and debonding in composites through noncontact laser ultrasonic scanning. *Compos Sci Technol* 100:10–18
- Rogge MD, Leckey CAC (2013) Characterization of impact damage in composite laminates using guided wavefield imaging and local wavenumber domain analysis. *Ultrasonics* 53:1217–1226
- Rose JL (1999) Ultrasonic waves in solid media. Cambridge University Press, Cambridge, UK
- Ruzzene M (2007) Frequency-wavenumber domain filtering for improved damage visualization. *Smart Mater Struct* 16:2116–2129

- Tian Z, Yu L (2014) Lamb wave frequency-wavenumber analysis and composition. *J Intell Mater Syst Struct* 25(9):1107–1123
- Tian Z, Yu L, Leckey C (2016) Rapid guided wave delamination detection and quantification in composites using global-local sensing. *Smart Mater Struct* 25:085042 (11pp)
- Wyatt RC (1972) Visualization of pulsed ultrasound using stroboscopic photoelasticity. *Non-Destr Test* 5(6):354–358
- Yu L, Tian Z, Leckey CAC (2015) Crack imaging and quantification in aluminum plates with guided wave wavenumber analysis methods. *Ultrasonics* 62:203–212



# Acoustic Microscopy

3

Frank Schubert, Martin Barth, Raffael Hipp, and Bernd Köhler

## Contents

Introduction to Scanning Acoustic Microscopy .....	76
Basics of Ultrasonic Wave Propagation in an Acoustic Microscope .....	78
Instrumentation of a Typical Acoustic Microscope for NDE Applications .....	84
Case Study: Scanning Acoustic Microscopy for Resistance Spot Weldings .....	87
Motivation .....	89
Ultrasonic Determination of the Weld Nugget Diameter .....	90
Ultrasonic Determination of the Thickness of the Weld Nugget .....	95
Advanced 3D SAM Imaging .....	99
3D Imaging in Scanning Acoustic Microscopy .....	101
General Focusing Techniques in Ultrasonic NDE .....	101
Test Specimen and X-Ray CT Image .....	103
Signal Preprocessing and Ultrasonic Reflection Tomography .....	103
Advanced SAM Imaging with Conical Annular Phased Arrays .....	108
Summary .....	111
References .....	112

### Abstract

The current chapter deals with scanning acoustic microscopy, a laboratory based high-precision ultrasonic NDE method for detailed investigations of the inner structure of solid materials and components. At the beginning, a short introduction into acoustic microscopy and its preferences is given. In part 2, the basics of ultrasonic wave propagation in an acoustic microscope are explained in terms of

---

F. Schubert (✉) · M. Barth · R. Hipp

Ultrasonic Sensors and Methods, Fraunhofer IKTS, Dresden, Germany

e-mail: [frank.schubert@ikts.fraunhofer.de](mailto:frank.schubert@ikts.fraunhofer.de); [martin.barth@ikts.fraunhofer.de](mailto:martin.barth@ikts.fraunhofer.de); [raffael.hipp@ikts.fraunhofer.de](mailto:raffael.hipp@ikts.fraunhofer.de)

B. Köhler

Accredited NDT Test Lab, Fraunhofer IKTS, Dresden, Germany

e-mail: [bernd.koehler@ikts.fraunhofer.de](mailto:bernd.koehler@ikts.fraunhofer.de)

wave front snapshots calculated by numerical simulations. In part 3, the typical instrumentation of a Scanning Acoustic Microscope (SAM) used in NDE is presented and some common applications are briefly discussed. In part 4, a comprehensive case study on SAM of resistance spot welding is presented, an exemplary and highly relevant application especially for the automotive industry. In the last two parts of this chapter, advanced imaging techniques for larger depths are presented, the first one based on a SAM adapted implementation of the 3D-SAFT algorithm (Synthetic Aperture Focusing Technique) and the second one using a specific transducer geometry with outstanding performance, a conical annular array.

---

## Introduction to Scanning Acoustic Microscopy

The acoustic microscope represents a tool to study the spatial variations of the elastic material properties of nontransparent solids. It therefore offers an insight into the internal microstructure of a material and potentially existing defects like cracks, pores, inclusions, and delaminations. Historical outlines of acoustic microscopy are highly interesting but are omitted here for reasons of a concise presentation. They can be found in several books and handbook articles (Briggs 1992, 1995; Gilmore 1999; Lemons and Quate 1979; Zinin 2001).

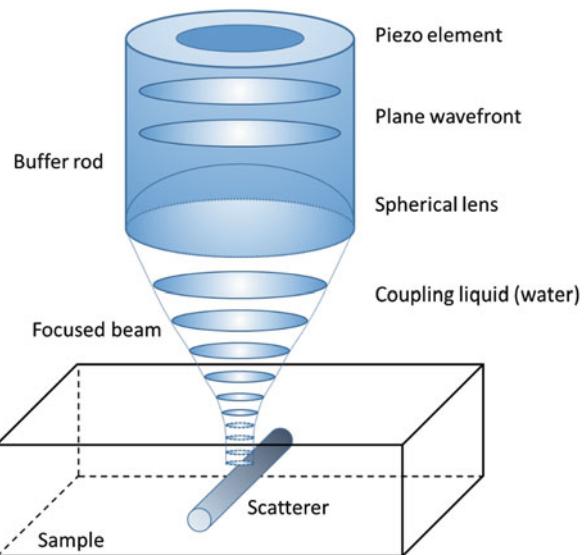
In scanning acoustic microscopy, a sample immersed in a coupling liquid (typically water) is imaged by ultrasonic waves excited and detected by piezoelectric transducers. The contrast in local reflection, produced by an acoustic impedance mismatch, yields a map of the spatial distribution of the mechanical properties. The data is collected by mechanically scanning the ultrasonic transducer across the sample surface and collecting the time-domain signal or A-Scan sequentially point by point. Finally, this data is further processed to produce C-Scan and B-Scan images by correlating time-of-flight information with sample depth based on a (usually very simple) wave speed model.

The typical operating frequencies of common ultrasonic microscopes used for nondestructive testing are between 10 and 400 MHz, resulting in wavelengths between approximately 7.5 and 300  $\mu\text{m}$ . For special applications where nearly optical resolution ( $\approx 1 \mu\text{m}$ ) is needed, systems up to 2 GHz are available. However, in this very high frequency range the wave attenuation in the coupling liquid is very strong (proportional to the frequency squared) and thus only very small working distances between transducer and sample are possible (e.g., approximately 50  $\mu\text{m}$  at 2 GHz). This article is focused on the lower frequencies with a higher relevance for typical NDE applications.

Scanning acoustic microscopes (SAMs) typically operate in reflection or pulse-echo mode (see Fig. 1). While in former times SAMs mostly used narrow band tone bursts for excitation, modern time-resolved systems use predominantly short broadband pulses even at very high frequencies. Therefore, a SAM can be seen as a very fast, accurate, and high-frequency ultrasonic scanner.

In order to improve the lateral resolution, focused wave fields are generally used. The corresponding focal spot is limited by diffraction and its lateral diameter is

**Fig. 1** Schematic diagram of reflection acoustic microscopy (inspired by Zinin and Weise 2003)



restricted to approximately half of the wavelength. This focusing can either be done by shaping the active transducer aperture to a spherical cap or by using a flat transducer together with an acoustic lens. The latter consists of a cylindrical buffer rod with a spherical cavity inserted on one end face of the cylinder (see Fig. 1). This face is often coated with a quarter-wavelength matching layer in order to increase transmission. The acoustic lens is mostly used in the high-frequency range above 70 MHz while spherically shaped transducers are dominating at lower frequencies.

Besides the spherical lens also cylindrical lenses with a line focus are often applied in NDE. They prove superior for the analysis of Rayleigh surface waves, especially in anisotropic materials. Depending on the specific application and the used wave form, the focal depth of the transducer can be changed by moving it mechanically up and down. However, a complete lateral scan is generally performed with a fixed focal depth.

In pulse-echo reflection microscopy, the same transducer emits the acoustic waves and also receives the signals scattered back from the sample. Since both emission and detection are focused and both focal points coincide, the SAM represents a confocal microscope. Since the transducer is placed in remote from the sample, no near-field effects of the waves can be deployed, classifying the SAM as a conventional far-field microscope (Zinin and Weise 2003).

Most SAMs today are also able to work in transmission mode. For this purpose, a second transducer is placed below the sample in order to detect the forward scattered waves. However, this arrangement is only possible if the sample is thin enough and the corresponding wave attenuation is sufficiently small. Another drawback of the transmission arrangement is the lack of time-of-flight information from the scatterer so that only its lateral position (but not its depth) can be determined. Transmission SAMs are often applied if the samples are very thin,

the material is only weakly scattering and also in cases where the material properties of anisotropic crystals (which typically show very low attenuation) need to be determined.

## Basics of Ultrasonic Wave Propagation in an Acoustic Microscope

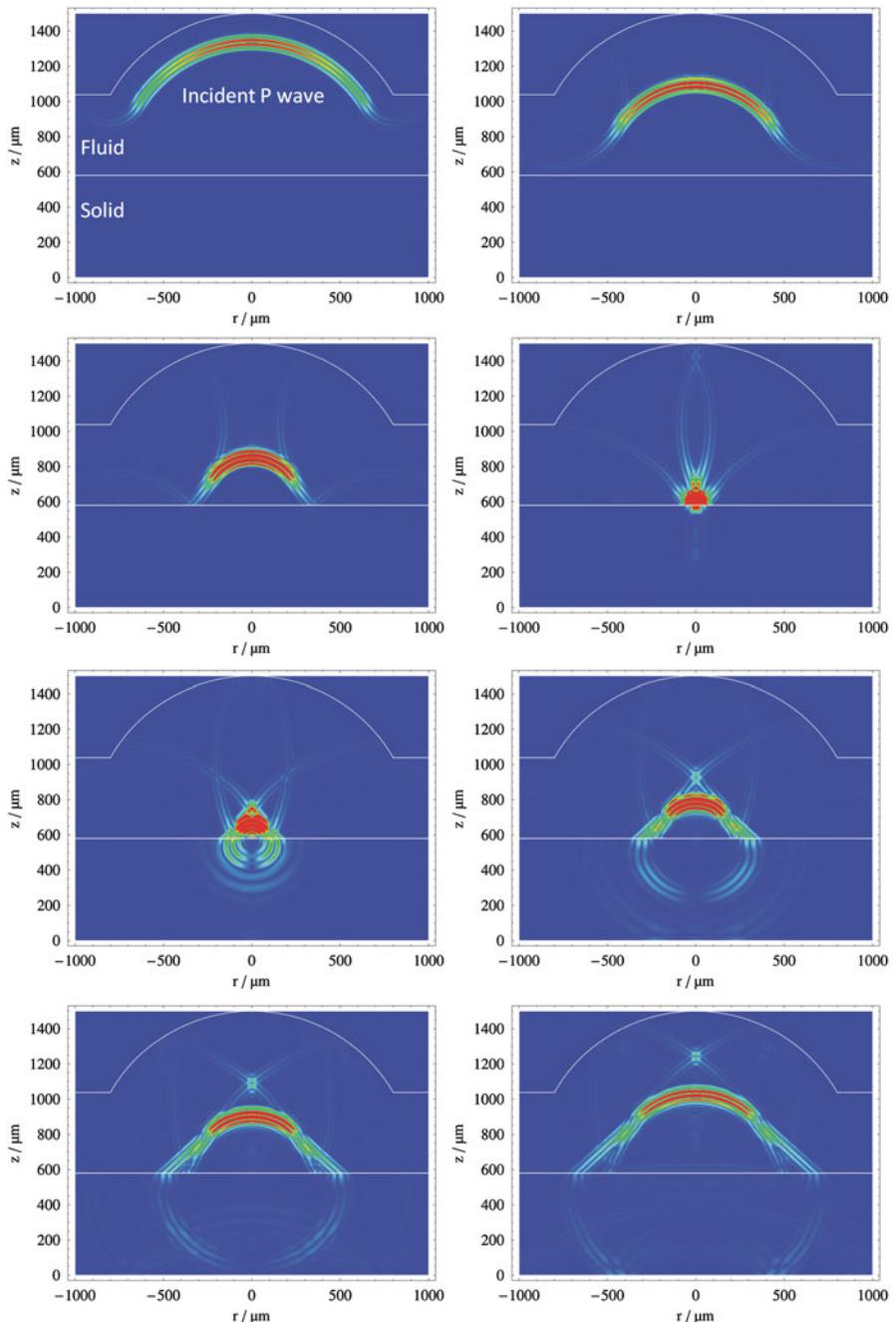
In order to demonstrate the basic wave modes relevant for data evaluation in an acoustic microscope, numerical simulations of ultrasonic wave propagation were performed based on the Elastodynamic Finite Integration Technique (EFIT), an explicit time-domain method for transient wave phenomena (Fellinger et al. 1995; Schubert et al. 1998, 2001, 2005). For this purpose, a simplified axisymmetric SAM model consisting of a spherically curved transducer and a fluid-solid interface was used (Figs. 2 and 3). The transducer with a lateral radius of 500  $\mu\text{m}$ , a (geometrical) focal length of 924  $\mu\text{m}$ , and a half-aperture angle of 60° excites a broadband RC2 pulse (raised cosine with 2 cycles) with a center frequency of 25 MHz.

The incident pressure or P wave, focused by the spherically curved aperture, propagates through the immersion liquid (water in this case) and hits the fluid-solid interface where it is strongly reflected due to the large acoustic impedance mismatch between the fluid and the solid (in this example,  $c_P = 4170 \text{ m/s}$ ,  $c_S = 2410 \text{ m/s}$ , and  $\rho = 7100 \text{ kg/m}^3$  were used). Moreover, several additional wave modes are generated as indicated in Fig. 3.

In the forward direction of the incident beam, a P wave and – by mode conversion – also a shear or S wave is generated. Both bulk waves show a significantly different directivity. In the solid, a Rayleigh surface wave moves along the fluid-solid interface. It produces a leaky wave (Zhu et al. 2004) at the other side of the interface which can be detected with special transducers. The strongest echo by far is the P-wave reflection from the fluid-solid interface but it is not the wave with the first arrival. A so-called precursor is geometrically built by the superposition of waves generated at the circular edge of the aperture, i.e., at the transition between curved and noncurved part of the transducer (compare snapshots in Fig. 2).

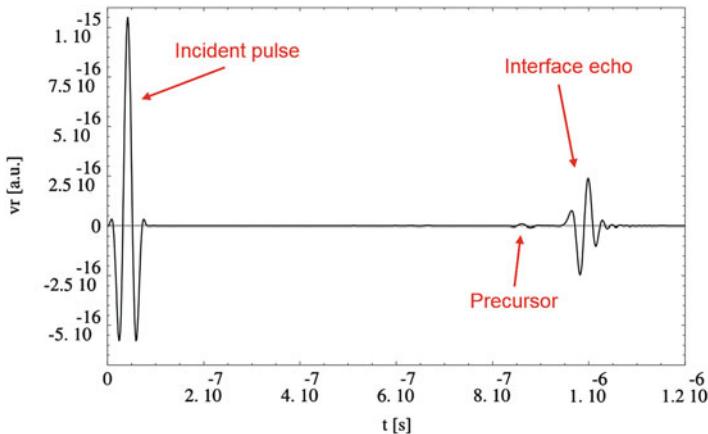
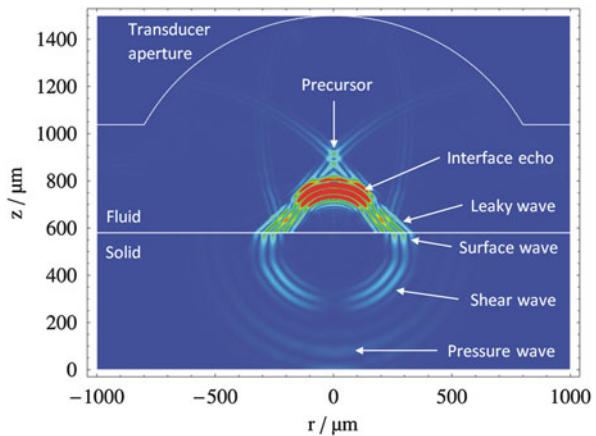
In Fig. 4, the resulting A-Scan for the model shown in Figs. 2 and 3 is shown. It incorporates the excitation signal, the reflected echo from the fluid-solid interface as well as the mentioned precursor. Due to the lack of defects in the solid it does not contain any indications from scatterers below the interface. These additional echoes, generated by the interaction of P, S, or Rayleigh waves with the defect, will occur after the arrival of the interface echo and their amplitudes will be much smaller. Therefore, in practice, the excitation signal and the interface echo are typically overamplified in order to let the defect echoes become visible (compare to section “Case Study: Scanning Acoustic Microscopy for Resistance Spot Weldings,” where a typical real-world A-Scan is shown).

In Fig. 5, two different examples for wave interaction with near-surface defects are depicted, a vertical 150  $\mu\text{m}$  long surface-breaking crack (cylindrical hole) whose lateral dimensions are small compared to the wavelength (left column) and a 100  $\mu\text{m}$



**Fig. 2** Snapshots of wave propagation in an ultrasonic microscope calculated by an axisymmetric version of the numerical EFIT method (Schubert et al. 1998, 2001)

**Fig. 3** Detailed view of the wave field with identification of wave modes

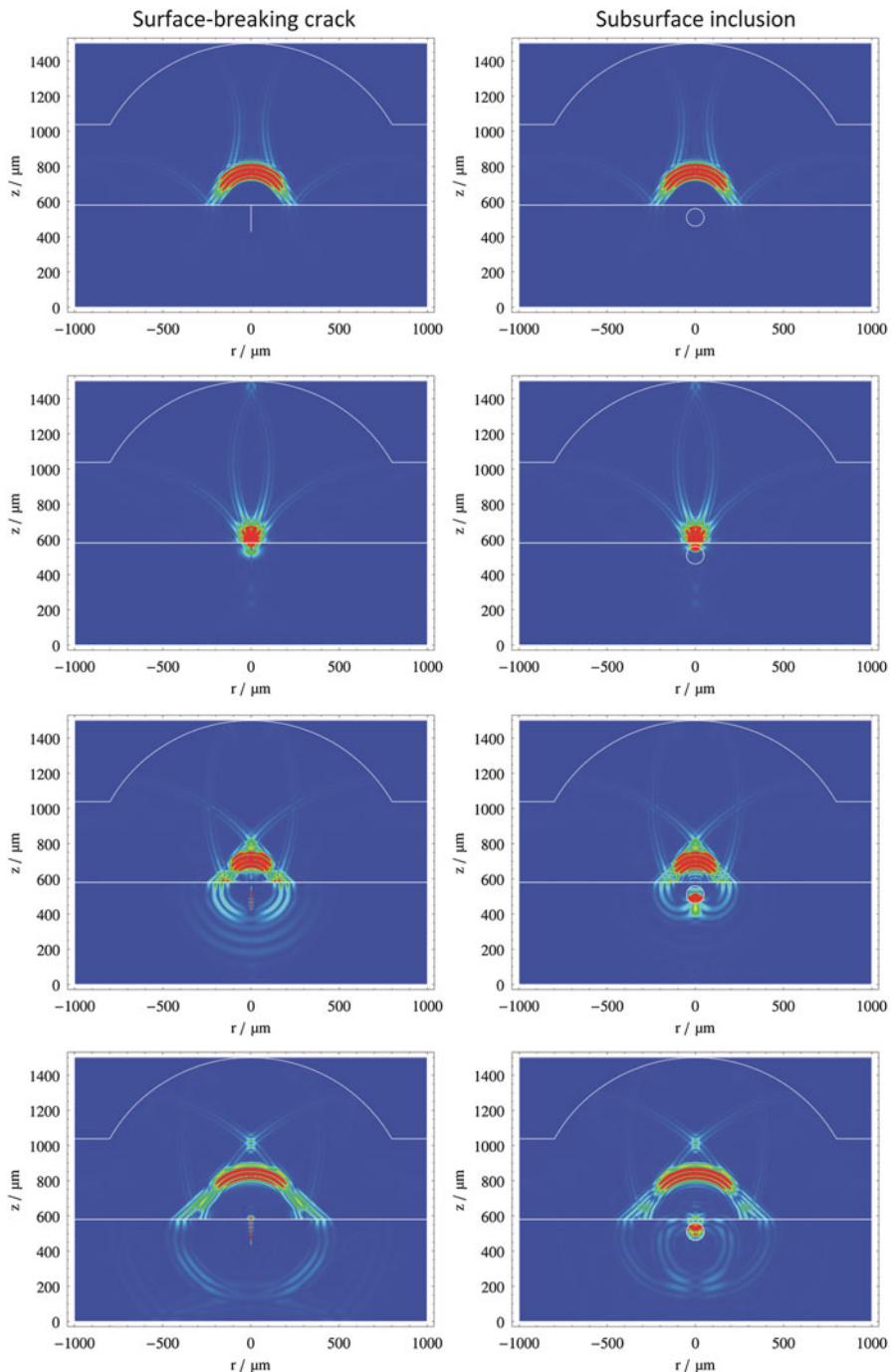


**Fig. 4** Calculated A-Scan of the particle velocity detected at the transducer aperture

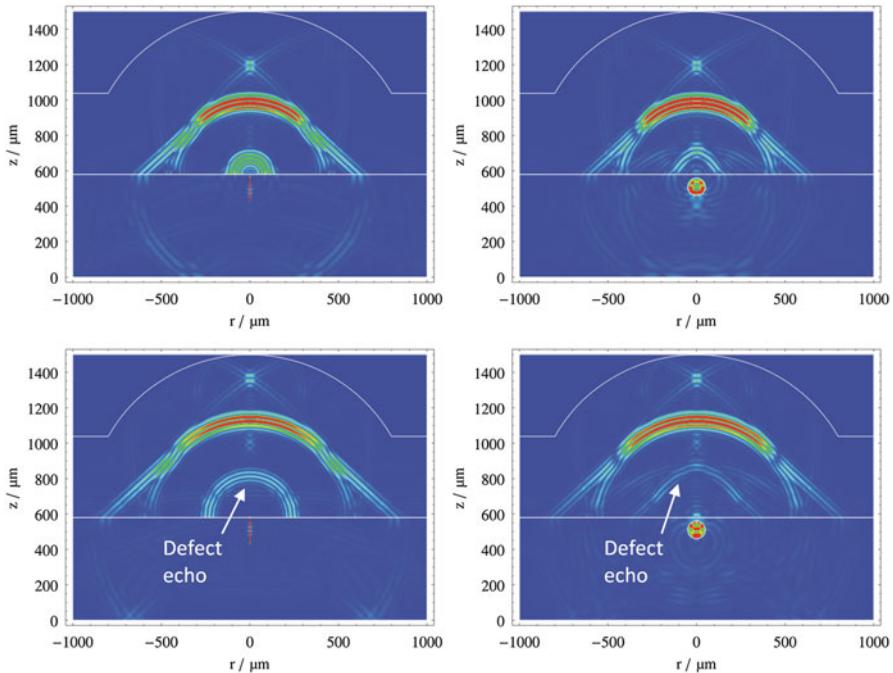
diameter spherical subsurface inclusion 70  $\mu\text{m}$  below the fluid-solid interface (right column). Both scatterers are filled with water and therefore show strong acoustic resonances.

In the first example, a strong P wave is generated inside the cylindrical crack, is reflected at its bottom tip, and is partly back-transferred to the immersion fluid. Other parts of the P wave are multiply reflected between top and bottom crack tip so that a periodic sequence of tip echoes with decreasing amplitudes arise (not shown here). In the case of the water filled inclusion, P waves are generated inside the inclusion and are multiply reflected between its top and bottom surface. Therefore, similar to the vertical crack described before, secondary echoes are transferred to the immersion liquid.

It is important to mention that the relative strengths of the wave modes shown in Figs. 3, 4, and 5 strongly depend on various factors like focal length,



**Fig. 5** (continued)



**Fig. 5** Snapshots of wave propagation in an ultrasonic microscope demonstrating interaction with near-surface defects. Left column: Cylindrical surface-breaking crack filled with water; right column: Spherical subsurface inclusion filled with water

half-aperture angle, and center frequency of the transducer, the material properties of the background medium and scatterer, the position of the focal spot relative to the fluid-solid interface, and also the location and depth of the defect.

In the present examples, the focal spot was always placed directly at the fluid-solid interface so that a quasi-point-like source with strong shear and surface waves was generated inside the solid. If the focal spot is placed below the interface (which is the common operation mode) the strength of these waves is reduced and the P wave becomes larger. In this case, however, the refraction of the ultrasonic beam at the interface must be taken into account. Since the wave speed in the solid is usually much larger than in the fluid, the focal length of the transducer in the solid is significantly shortened compared to the free wave field in the fluid according to Snell's law.

Due to the spherically shaped form of commercially available acoustic lenses (instead of aspherical geometry), distant rays are differently refracted at the lens-fluid interface than rays close to the axis. This effect is further increased by the limited aperture angle of real transducers. Therefore, the focal region of a typical transducer used for SAM measurements is effectively a stripe along the

longitudinal axis rather than a point. This means that the axial resolution of a SAM is usually significantly worse than the lateral resolution.

According to the so-called Rayleigh distance  $z_{\text{lateral}} = 0.61\lambda/\sin \alpha$ , which determines the lateral resolution, and the axial resolution  $z_{\text{axial}} = 0.5\lambda/(1 - \cos \alpha)$  (see Zinin and Weise 2003), we obtain for the ratio of axial to lateral resolution  $r = 0.82 \sin \alpha/(1 - \cos \alpha)$  with the semi-aperture angle  $\alpha$ . This value no longer depends on the wavelengths  $\lambda$  in the immersion fluid and ranges from  $r = 9.4$  for  $\alpha = 10^\circ$  to  $r = 1.4$  for  $\alpha = 60^\circ$ .

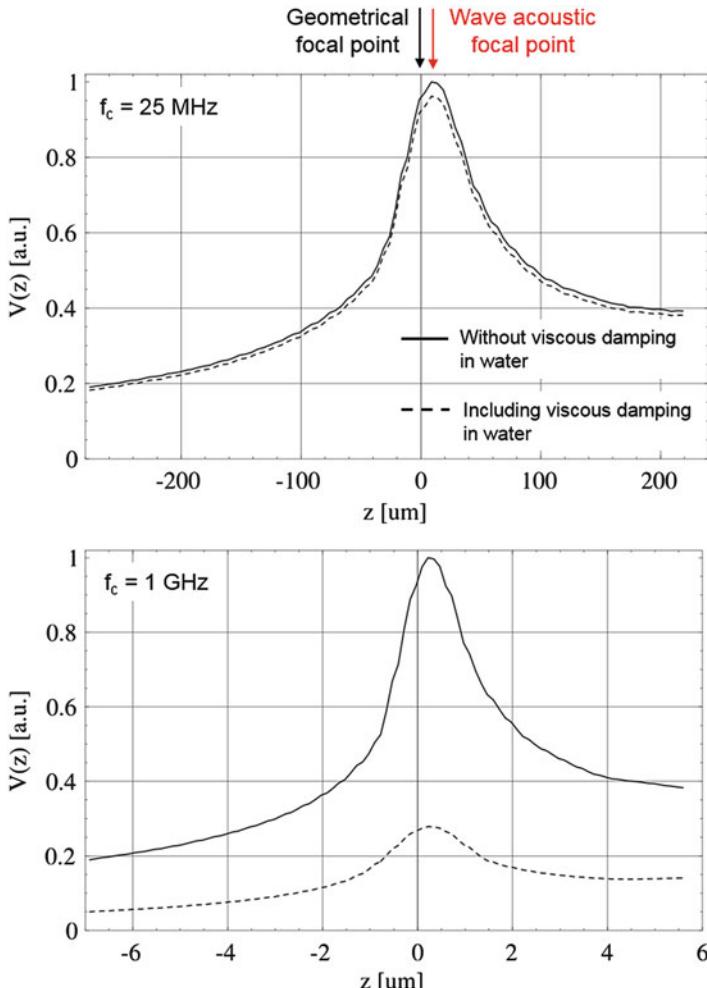
In this context, the attenuation of the different frequencies in the immersion liquid is of utmost importance because it directly affects the usable wavelengths and thus, the reachable axial and lateral resolution of the SAM. This effect can be demonstrated by calculating so-called  $V(z)$  curves, i.e., reflection amplitudes as a function of the distance between geometrical focal point and scatterer. In Fig. 6,  $V(z)$  curves for an ideally reflecting motion-free and flat boundary are shown for different center frequencies of  $f_C = 25$  MHz (top picture) and  $f_C = 1$  GHz (bottom picture).

In the figure  $z = 0$  means that the boundary is placed exactly at the geometrical focal plane. Negative  $z$ -values indicate a defocusing below the boundary, positive  $z$ -values a defocusing above the boundary. As can be seen from the maxima of the  $V(z)$  curves, the wave acoustic focal point does not exactly correspond to the geometrical one. Instead it is shifted by 10  $\mu\text{m}$  (for  $f_C = 25$  MHz) and 0.25  $\mu\text{m}$  (for  $f_C = 1$  GHz), respectively, to positive  $z$ -values. This shift amounts to 1% of focal length and 1/6 of the wavelength in water.

The 25 MHz simulation shown in Fig. 6 (top) is based on an EFIT model in which viscosity of water is neglected. In this context, a simulation at 1 GHz leads to the same  $V(z)$  curve because the 1 GHz model can simply be treated as a linearly rescaled version of the 25 MHz model (with higher frequency but correspondingly smaller dimensions, see solid line in Fig. 6, bottom). However, with EFIT it is also possible to incorporate viscous damping in the model. In the case under consideration this has been done by using a viscosity of  $\eta = 3.08 \times 10^{-3}$  Pa s for water at room temperature. The results of these additional calculations are represented by the dashed lines in Fig. 6.

One can see that at 25 MHz, viscosity in water only slightly decreases the amplitude of the  $V(z)$  curve. However, at 1 GHz the effect is significantly stronger. The amplitude is decreased to 28% of the original value. Moreover, the width of the  $V(z)$  curve is increased which means that the focal point is less sharp than in the case without damping. The described phenomena are based on the fact that viscous damping is directly proportional to the square of frequency and thus, higher frequencies are attenuated much stronger than lower ones. Thus, simple linear rescaling of models with different frequencies is no longer possible.

The results clearly demonstrate that viscous damping in water can only be neglected if relatively low frequencies are used. For frequencies above 30–40 MHz, the amplitude decay due to viscous damping is larger than 10% and can no longer be omitted. In all preceding examples (Figs. 2–5), center frequencies of 25 MHz were used and viscosity was neglected.



**Fig. 6**  $V(z)$  curves of a motion-free flat boundary obtained at center frequencies of 25 MHz (top picture) and 1 GHz, respectively (bottom picture). The solid curves belong to a numerical EFIT model without viscous damping in water, and the dashed curves indicate a model with damping taken into account

## Instrumentation of a Typical Acoustic Microscope for NDE Applications

A typical acoustic microscope for NDE applications is shown in Fig. 7. It consists of four main components:

- A fast and accurate mechanical x-y-z scanner with dynamic linear motor drive for immersion testing in a built-in water tank.



**Fig. 7** Scanning acoustic microscope for NDE applications with mechanical high-speed scanner and ultrasonic hardware located on the left and the PC system for data evaluation on the right (courtesy by PVA Tepla Analytical Systems GmbH, Germany)

- A set of ultrasonic transducers with varying center frequency, focal length, and semi-aperture angle for different applications and wave modes.
- A powerful ultrasonic electronics for high-frequency data acquisition and signal processing.
- A PC with multiple screens and SAM software for data storage, data evaluation, and imaging of the results.

The nominal frequency range of the SAM shown in the figure lies between 5 and 400 MHz, the most commonly used frequencies, however, are between 10 and 200 MHz. With the integrated water tank, samples up to a size of  $300 \times 300 \times 100 \text{ mm}^3$  can be investigated. For significantly smaller or larger samples, other systems with adapted dimensions of scanner and water tank are available.

One of the most frequent applications of such a SAM is the testing of wafers with microelectronic components like integrated circuits (see Figs. 8 and 9).

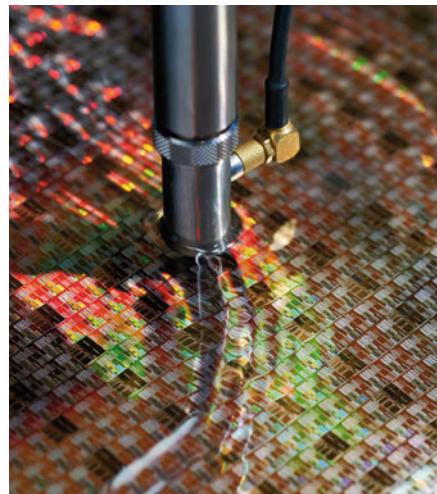
Besides microelectronics, a variety of other materials and components can be tested with an ultrasonic microscope, e.g., single crystals, multi-layered structures, adhesive layers, polycrystalline materials, fiber composites, and particulate composites, to name only a few. For this purpose, various ultrasonic transducers with different geometries and frequencies are available. Figure 10 shows a typical set of ultrasonic probes used for NDE applications in the frequency range between 10 and 200 MHz.

As mentioned earlier, different types of transducer technologies are applied for different frequencies. In the low frequency range between 5 and 20 MHz, PZT based transducers with spherically curved apertures are used (Fig. 11, left hand side). In the intermediate range between approximately 30 and 75 MHz, curved transducers



**Fig. 8** Detailed view of water tank and scanner unit (on the left) with ultrasonic transducer currently scanning a wafer with microelectronic components

**Fig. 9** High-frequency ultrasonic probe during wafer scan. The transducer is partly immersed in the water in order to provide good and stable coupling conditions and to avoid heavy wave movement



based on PVDF films are efficient (Fig. 11, center). In the high frequency range above 75 MHz, flat thin film transducers with a sapphire buffer rod and an integrated acoustic lens are dominant (Fig. 11, right hand side).

In Fig. 12, the Graphical User Interface (GUI) of a SAM software (top picture) is shown together with a typical A-Scan (bottom picture). The latter serves as a basis for B- and C-Scans using specific time gates and amplitude thresholds.

As mentioned before, a typical and common application for SAM systems as shown in Figs. 7 and 8 is the investigation of microelectronic components. In the top



**Fig. 10** Set of ultrasonic transducers for acoustic microscopy



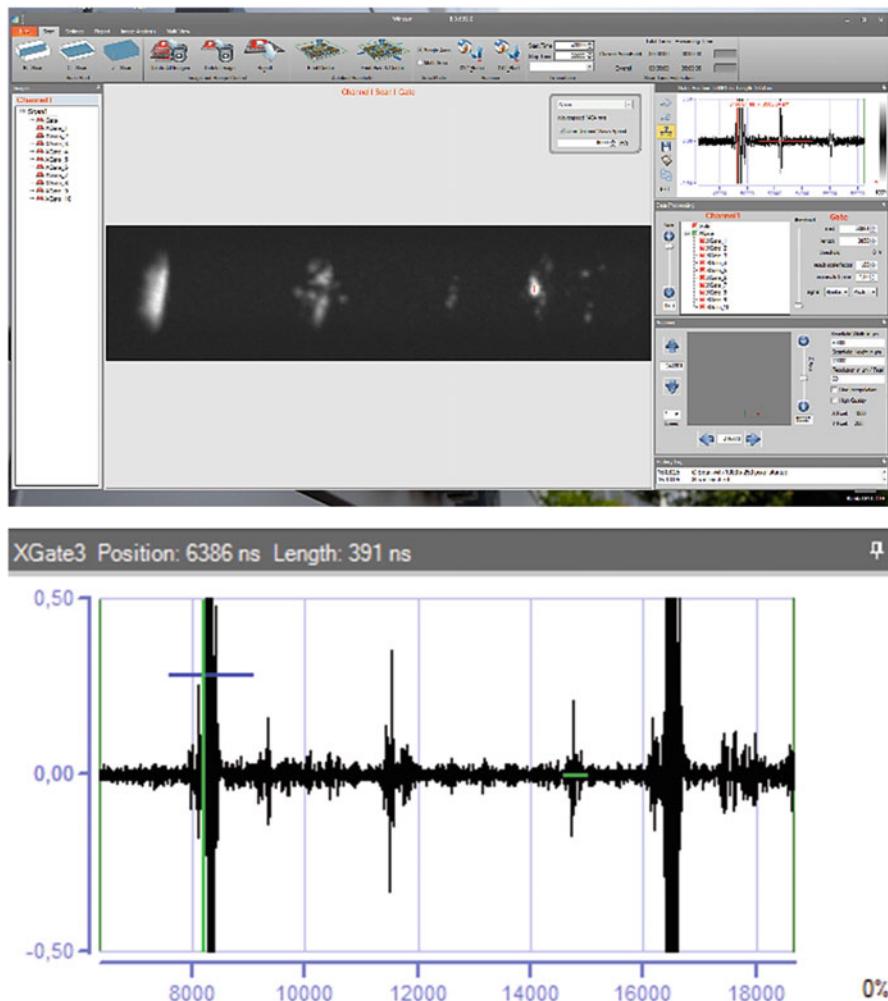
**Fig. 11** Ultrasonic probes for the low (on the left), intermediate (center) and high frequency range (on the right)

row of Fig. 13 various C-Scans obtained at different depths of an electronic chip are shown. In the second row, a number of other materials and applications are depicted, including polycrystalline silicon, potato cell structures, human prostate tissue, and CFRP T-fitting measured from the base plate. All scans in Fig. 13 were performed in the low frequency range between 10 and 30 MHz.

---

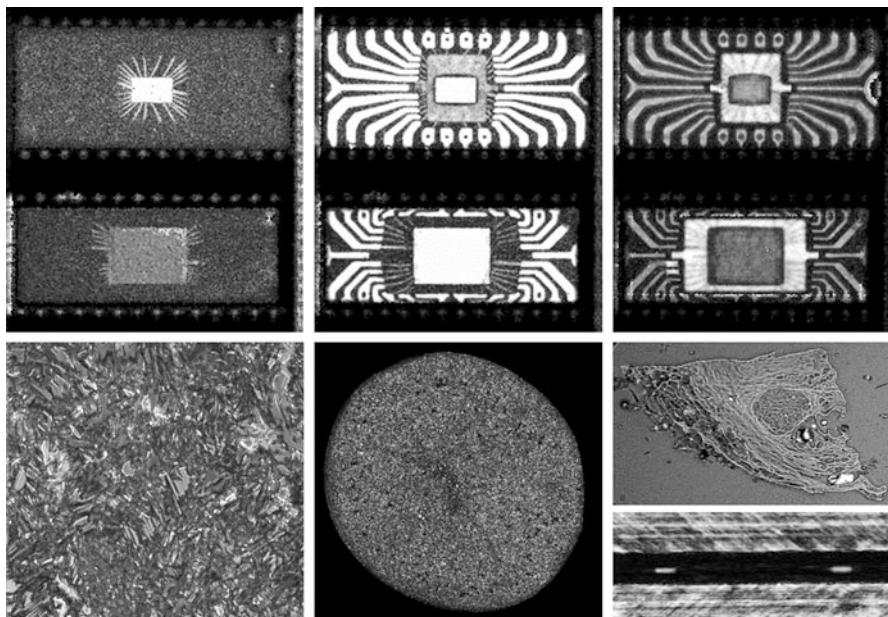
### Case Study: Scanning Acoustic Microscopy for Resistance Spot Weldings

For nondestructive testing of resistance spot weldings, several commercial and mobile ultrasonic NDT systems are available. Most of them offer no image at all or only a very limited C-Scan image resolution at the utmost. With scanning



**Fig. 12** Graphical user interface of a SAM software (top) and a typical A-Scan obtained at a single measuring point (bottom) as basis for B- and C-Scan imaging

acoustic microscopy, it is possible to obtain high-resolution B- and C-Scans and to perform additional quantitative analysis of high-frequency A-Scans. This allows for a sophisticated analysis of the weld nugget. For instance, it is possible to determine not only the lateral dimension (diameter) of the nugget but also its approximate thickness. The latter is obtained by analyzing the effective ultrasound attenuation caused by the interaction with the modified grain structure inside the nugget. Therefore, scanning acoustic microscopy can be used as a quantitative reference and calibration tool for manual spot-weld testing systems (Schubert et al. 2014).

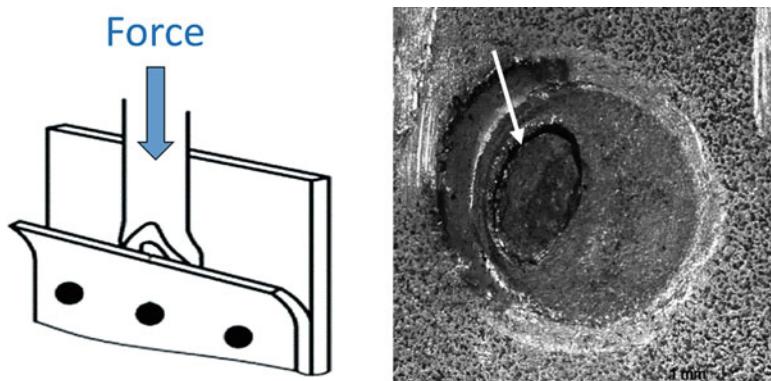


**Fig. 13** Typical applications in scanning acoustic microscopy, depicted by C-Scan images. Top row: Microelectronic chip structure at various depths; bottom row: grains of polycrystalline silicon (left), cell structure of a potato slice (center), human prostate tissue (top right), and delaminations and fiber bundles in a CFRP T-fitting (bottom right)

## Motivation

Resistance spot welding is an established joining technology in the automotive industry, in frame-and-body construction, and in sheet metal forming. It is characterized by a high cost-effectiveness and process reliability. The quality of spot welds can be determined by destructive and nondestructive testing methods as well as by an indirect analysis of the process parameters. In contrast to a fuzzy parameter analysis, destructive and nondestructive techniques allow for a direct quantitative evaluation of the spot weld. In the laboratory, the weld quality is often determined by a chisel test. During this procedure, the chisel is mechanically forced between the two metal sheets until one of the sheets is removed by unbuttoning (Fig. 14, left picture). After that the unbuttoned area (marked by a white arrow in the picture on the right) can be investigated by optical microscopy in order to determine geometry, diameter, and type of fracture.

During production, however, NDT methods are essential. The most important NDT method for resistance spot welds is ultrasonic testing. Besides single-channel systems with integral analysis of the back-wall echo, multichannel matrix systems and a miniature scanner with combined translation/rotation are available as well. In the present work, the RSWA matrix system of Tessonics Inc., Canada (Maev et al. 2004; Maev et al. 2011) was evaluated. A scanning acoustic microscope Evolution II of PVA Tepla Analytical Systems GmbH (Germany) served as reference for high-



**Fig. 14** Destructive chisel test of a resistance spot weld (on the left). After one of the metal sheets has been lifted, the unbuttoned area (white arrow in the picture on the right) can be further analyzed with respect to geometry, diameter, and type of fracture

precision investigations. For the first part of this study as described in subsection “[Ultrasonic Determination of the Weld Nugget Diameter](#),” two sheet combinations with varying sheet thicknesses and materials were used. The upper sheet had a thickness of 0.65 mm in each case and was made of a deep drawing steel DX56 + Z100MB. The lower sheet consisted of a corrosion-resistant steel X5CrNi10–18 with a thickness of 2 mm and a deep-drawing steel DX56 + Z100MB with thicknesses of 1 and 2 mm, respectively. All sheets were zinc-plated.

### **Ultrasonic Determination of the Weld Nugget Diameter**

Ultrasonic testing of resistance spot welds is usually based on pulse-echo measurements and subsequent analysis of A-, B-, or C-Scans.

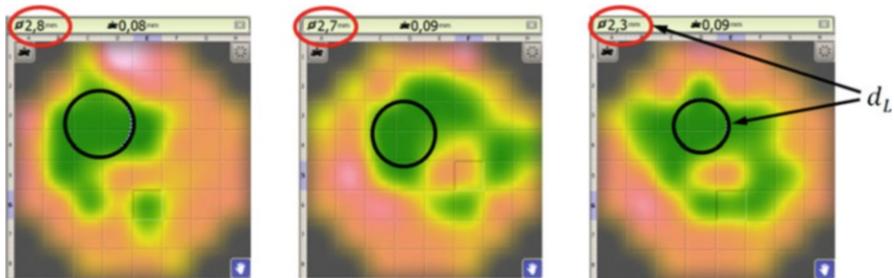
### **Manual Testing**

The RSWA system of Tessonics Inc. Canada, is a mobile ultrasonic testing device for manual industrial inspection (Fig. 15). It is based on an  $8 \times 8$  matrix transducer with 52 piezoelectric elements. Each element is acting separately in pulse-echo mode (no phased array) using a frequency of approximately 15 MHz. Due to the circular shape of a spot weld, the three elements in the four corners of the matrix are omitted which yields the final number of 52 active elements. The full aperture of the transducer amounts to approximately  $10 \times 10 \text{ mm}^2$  and is protected by a polystyrene delay line that is coupled to the specimen by a conventional coupling paste.

For a complete amplitude C-Scan of the area under test, the A-Scans of all 52 elements are processed and evaluated separately. The resulting picture is interpolated in order to obtain a smoother gradient in the final C-Scan. The amplitudes are displayed by a color scale in which green colors indicate a properly welded and red colors a non-welded region (Fig. 16).



**Fig. 15** RSWA Ultrasonic system from Tessonics Inc. (picture on the left) and corresponding matrix array transducer (picture in the center and on the right)



**Fig. 16** Typical results of the RSWA system showing welded (green) and nonwelded regions (red) inside the matrix aperture. In this case, the measurement across a single spot weld was repeated three times in order to demonstrate the statistical fluctuations due to varying coupling conditions

Inside the C-Scans, a square grid structure is visible according to the 52 single elements of the matrix transducer. For a quantitative evaluation of the weld nugget, a black circle is automatically placed inside the C-Scan. It represents the largest circle that lies completely inside the green area (welded region). The resulting effective diameter of the nugget,  $d_L$ , is given in the top left corner of the displayed image. Additionally, the depth of indentation of the upper welding electrode is shown in the middle of the same bar.

In general the RSWA system, designed for economic practical use, delivers accurate results of spot weld quality in case of good surface conditions. Due to the hard delay line, the results are sometimes affected by tilt effects. In this case, multiple measurements across one and the same spot weld could lead to different results of weld quality as demonstrated in Fig. 16. According to DIN EN ISO 14327,  $d_L$  has to be larger than 2.45 mm in order to guarantee a good weld quality for this specific metal sheet combination. In two of the three measurements this criterion is fulfilled but in the third measurement  $d_L$  is smaller than 2.45 mm indicating a nonacceptable weld that would be rejected. Besides the problem of reproducibility the RSWA system shows several advantages like high mobility, easy operation, and a short measurement time of approximately 3 s per spot weld.

## SAM Inspection

In order to evaluate the results of the RSWA system for manual testing, the SAM was used as a laboratory reference device (see Figs. 7 and 8). It consists of a fast and precise scan unit placed above an immersion tank in which the spot-weld specimen is situated. Deionized water is used as coupling medium in order to guarantee stable and reproducible coupling conditions.

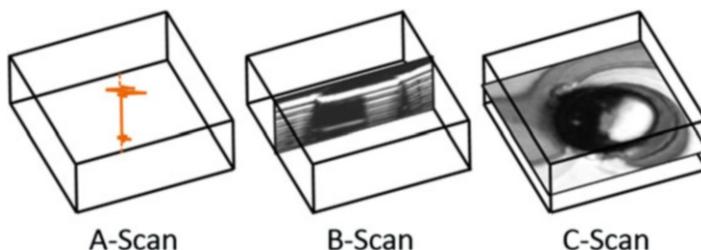
The mechanical scan unit can be moved along three separate axes, x, y, and z. For each scan point, a pulse-echo measurement is performed. Usually the whole area of interest is scanned across the x/y-plane by using a fixed distance between probe and surface according to the specific focal length of the transducer and the expected depth of the defect. By using various transducers the testing frequency can be adapted to the current testing configuration. In this example, the same frequency as for the RSWA system (15 MHz) was used. For evaluation of the SAM data, conventional A-, B-, and C-Scans were considered (Fig. 17).

In Fig. 18 (left picture), the typical C-Scan of a bad quality spot weld is depicted. The grey-scale indicates the amplitude of the interface echo between the two metal sheets. The size of the scan area is identical to those used by the RSWA system (compare Fig. 16). In the C-Scan, three different weld qualities can be identified, i.e., good welding (dark area 2, no or weak interface echo), no welding (light area 1, strong echo from the air gap between the metal sheets), and a so-called weld bond where only the zinc layers of the sheets joined together (small interface echo, indicated by 3).

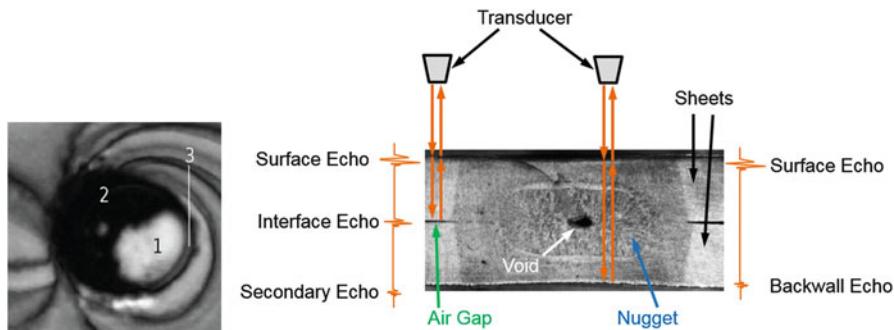
Moreover, a void can also be identified as a small light spot below number 2. In Fig. 18 (picture on the right), the possible travel paths and types of echoes of ultrasonic waves in regions 1 and 2 of a spot weld are displayed. In the following, the three characteristic regions of the spot weld are described in detail by analyzing the corresponding A- and B-Scans.

Non-welded region 1:

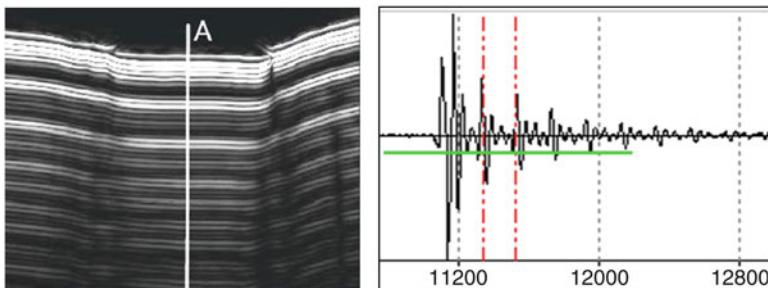
If the two metal sheets are not welded, an air gap between them remains which leads to a strong interface echo or more precisely, to a multiple echo from the back wall of the upper sheet. Nearly no energy is transmitted to the lower sheet and thus, no echo from its back wall can be identified (Fig. 19).



**Fig. 17** Conventional scan modes used for SAM inspection of resistance spot welds



**Fig. 18** Principle of ultrasonic microscopy of a spot weld. On the left: C-Scan with three different weld regions. On the right: Relevant types of echoes for welded and nonwelded regions



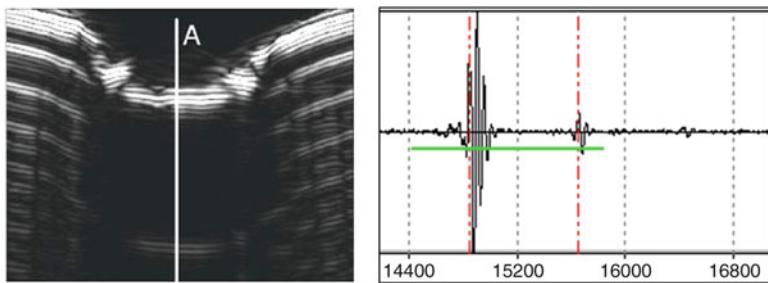
**Fig. 19** B-Scan (on the left) and A-Scan (on the right) for a nonwelded region of the spot weld. The temporal distance between the multiple echoes correlates to the thickness of the upper metal sheet. The A-Scan on the right was measured at position A as indicated in the B-Scan in the left picture

#### Welded region 2:

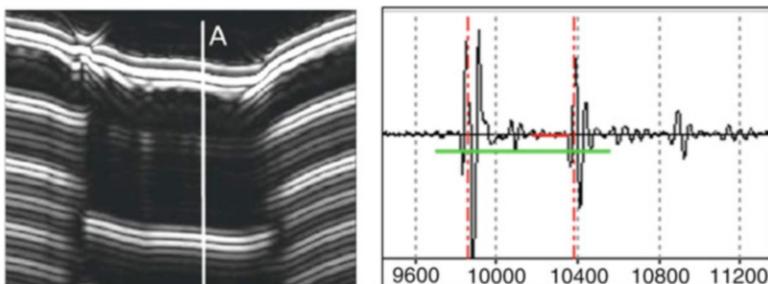
In this case, the two metal sheets were completely welded and thus, only the surface echo and the (weak) back wall echo from the lower sheet are visible. Their temporal distance correlates to the thickness of the complete sheet combination. In contrast to region 1, no interface echo can be identified (Fig. 20).

#### Weld-bonded region 3:

If the welding process only leads to a fusion and bonding of the zinc layers (with lower melting temperature than the base material) a so-called weld bond is generated. In this case, an air gap no longer exists but an additional interface echo occurs even if the two metal sheets are made of the same material (Fig. 21). It is evident that the back wall echo from the lower sheet is stronger than the corresponding echo from the welded region 2. A possible explanation is based on the fact that in case of a weld bond, no microstructural



**Fig. 20** B-Scan (on the left) and A-Scan (on the right) for a properly welded region of the spot weld. No interface echo is visible. Instead, the back wall echo from the lower metal sheet appears. However, in case of two different sheet materials, a weak interface echo might be expected due to the impedance mismatch



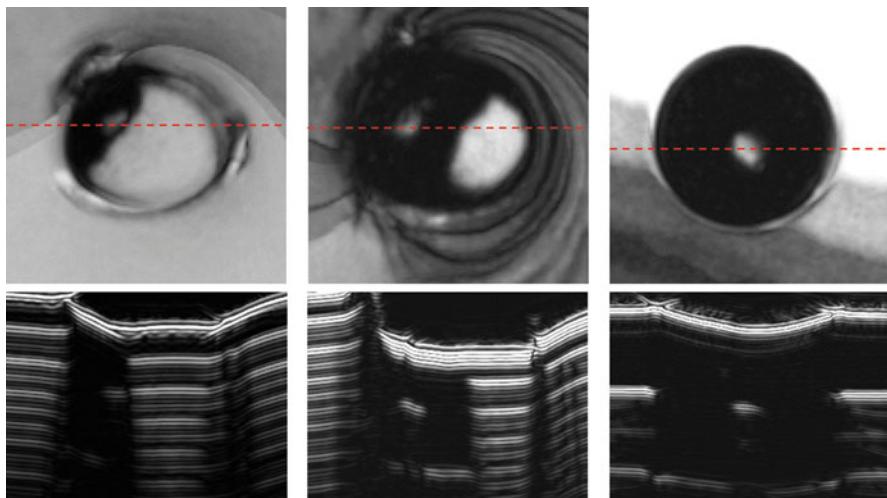
**Fig. 21** B-Scan (on the left) and A-Scan (on the right) for a weld-bonded region of the spot weld. Compared to the welded region 2, an additional small interface echo and a stronger back wall echo from the lower sheet occur

transformation takes place and thus no additional ultrasound attenuation due to scattering exists.

In Fig. 22, three further examples of defective spot welds including nonwelded regions and an additional void are shown. In addition to the C-Scan in the upper row, characteristic B-Scans along the red-dotted lines are depicted as well. They give a deep insight into the details of wave propagation.

It is worth mentioning that all three spot welds in Fig. 22 are characterized by a distinct surface mark of the welding tongs, but the wave propagation is not heavily affected so that a clear discrimination between welded and nonwelded parts is still possible.

In Fig. 23, further SAM C-Scans of spot welds are shown and compared with the results of the RSWA system revealing the significantly better resolution of the acoustic microscope. Such a comparison can be very helpful to identify the limitations of the current manual testing systems and to further improve them.



**Fig. 22** Results of SAM inspection of three different spot welds with defects. Upper row: C-Scans of the midplane of the sheet combination; bottom row: B-Scans obtained along the red-dotted lines

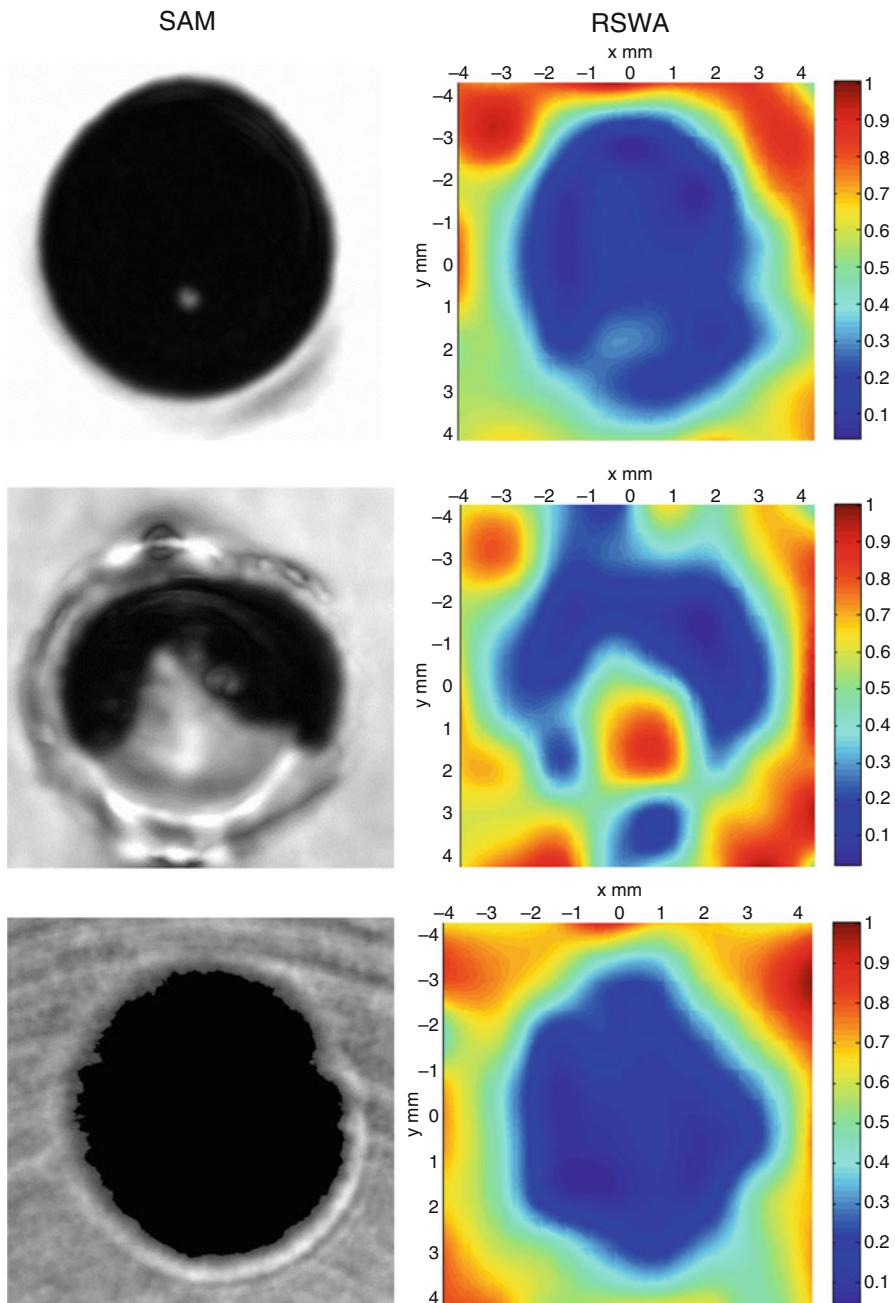
### Ultrasonic Determination of the Thickness of the Weld Nugget

A deeper analysis of the results shown before raises the question if it is also possible to extract the thickness of the weld nugget based on the ultrasonic attenuation of the back wall echo. In case of the bonded region shown in Fig. 20, the back wall echo of the lower metal sheet is significantly weaker than in Fig. 21 where the weld-bonded case is displayed. In case of a good welding, the grain structure of the welded zone is modified as can be seen in Fig. 18 (on the right) for instance. Due to the larger grains, the attenuation caused by scattering is stronger and thus, the back wall echo becomes smaller. Therefore, if the total thickness of the metal sheet combination is known, it should be possible to determine the approximate lateral dimension of the weld nugget, i.e., its thickness, from the analysis of the back wall echo.

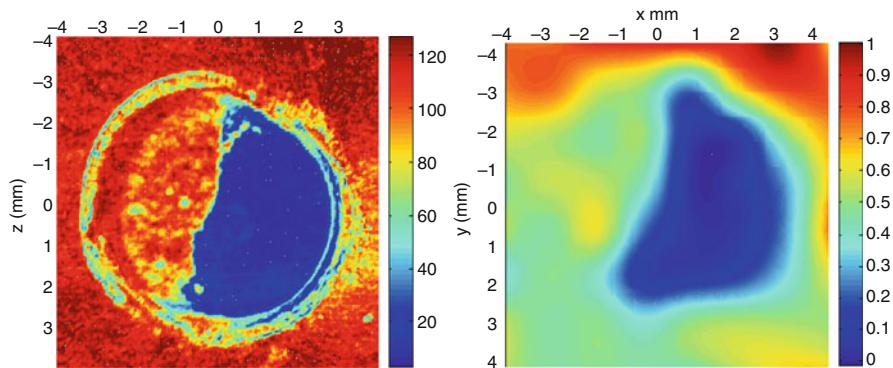
### Effect of Surface Topography

While investigating the ultrasonic attenuation caused by microstructural changes during the welding process one has to be sure that no other effects mask the measurements. The first goal therefore was to verify if the surface topography and the inclination of the upper surface shows a significant influence on the amplitude of the back wall echo.

For the new investigations two different metal sheet combinations were used. In combination A, both the upper and the lower sheet were 1.5 mm thick and made of uncoated DC01 material (a cold-rolled low carbon steel). In combination B, the upper sheet had a thickness of 0.8 mm and was made of DX56 + Z100MB (zinc-plated) while the lower sheet was 1.0 mm thick and made of the same material. In a first step, the surface topography of each spot weld was measured by extracting the



**Fig. 23** (continued)



**Fig. 23** C-Scans obtained by a high-resolution scanning acoustic microscope (left column) and by the RSWA system for manual inspection (right column). The first three SAM pictures (grey-scale) were obtained at 15 MHz, i.e., the same frequency as used in the RSWA system. The last colored SAM picture was measured at 50 MHz

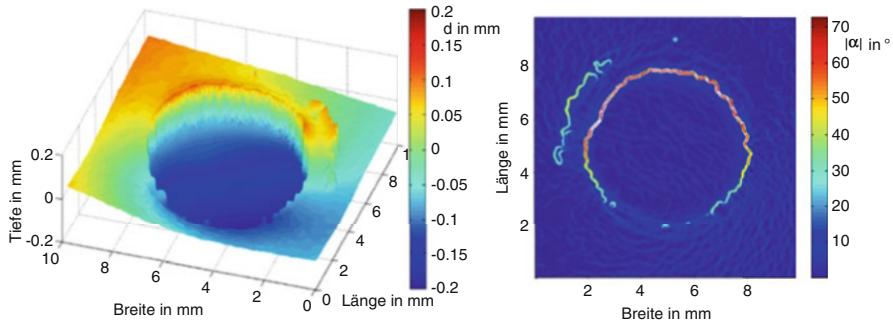
time-of-flight of the surface echo and therewith, the distance of each surface point to the US transducer (see Fig. 24, picture on the left). From this topography, an inclination angle map was determined which shows the absolute value of the maximum inclination angle in a linear color scale (Fig. 24, picture on the right).

From Fig. 24, it is evident that ultrasonic testing is only critical at the edges of the electrode indentation where the inclination angle is very high. However, in the inner circular indentation area the surface is sufficiently smooth. Moreover, the overall inclination of this inscribed circle is always below 5°. A couple of simulation studies based on acoustic ray tracing revealed that such a tilt angle has no significant influence on the amplitude of the back wall echo. In order to verify this theoretical finding, the surface and back wall echoes from all available specimens with a good welding were first measured without a prior surface treatment. After that the surface was mechanically treated by face milling in order to obtain a perfectly plane and smooth surface. Then the measurements were repeated and compared with the initial measurements without surface treatment. The results are given in Fig. 25 for both metal sheet combinations.

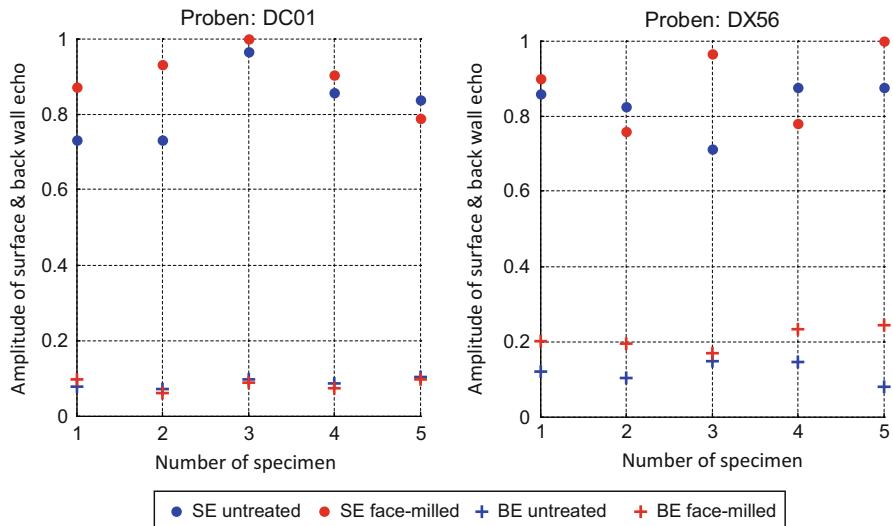
From Fig. 25, one can see that the differences between mechanically treated and untreated surfaces are rather small and lie within the range of only a few percent, especially for the back wall echo. Therefore, it can be concluded that the effects of surface roughness and inclination can be neglected and no mechanical surface treatment is necessary, at least for this kind of immersion testing performed in the SAM.

### Effect of Grain Structure

After excluding the possible influences of surface and coupling conditions, the effect of the grain structure itself was investigated. From Fig. 18 (on the right), it is obvious that the grain structure within the weld nugget is different from the microstructure in the surrounding heat-affected zone. A typical color encoded C-Scan of the back wall echo is shown in Fig. 26.



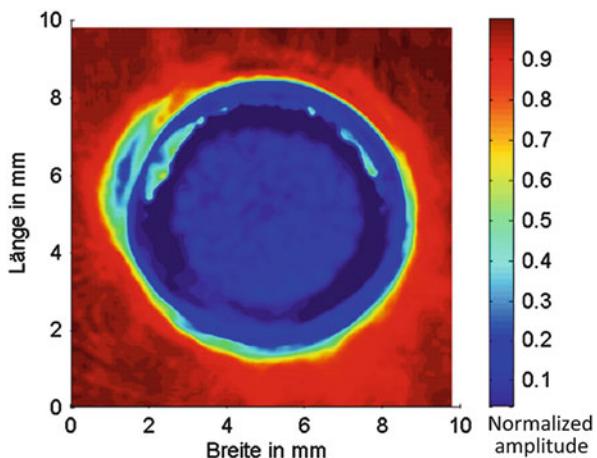
**Fig. 24** Typical surface topography (on the left) and inclination angle map (on the right) extracted from the time-of-flight data of the upper sheet surface echo



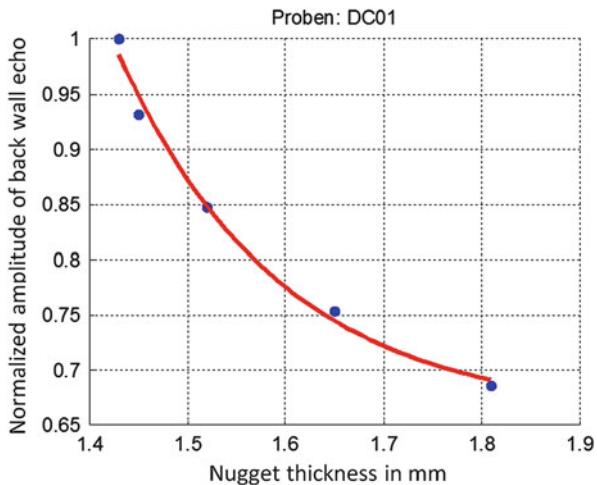
**Fig. 25** Measurement results for the normalized amplitude of the surface echo (SE,•) and the back wall echo of the lower sheet (BE,+) for untreated (blue) and face-milled surfaces (red) for five specimens of sheet combination A (DC01, on the left) and B (DX56, on the right)

The red parts in Fig. 26 indicate nonwelded metal sheets with a strong interface echo. The dark blue regions coincide with the “crater rim” of the electrode indentation with no back wall echo. The inner light-blue circle indicates the welded part and is of particular interest since it shows a heterogeneous character caused by the grain microstructure and contains information about the traversed thickness of the weld nugget. In order to demonstrate the latter aspect, the lateral dimension of the nugget was extracted destructively by optical micrographs after the ultrasonic measurements. Based on this data the echoes of the back wall echo of the lower sheet could be drawn as a function of the thickness of the weld nugget as shown in Fig. 27.

**Fig. 26** Color encoded normalized C-Scan of the back wall echo of the lower sheet



**Fig. 27** Normalized amplitude of the back wall echo of the lower sheet as a function of weld nugget thickness for sheet combination A (DC01). The measurements were performed at a scan point in the center of the weld nugget in each case



In case of the DC01 sheet combination, the experimental data very precisely follow an exponential function with a damping coefficient of  $\gamma_{DC01} \approx 6060 \text{ m}^{-1}$ . For practical use, the curve shown in Fig. 27 can be calibrated by one or two destructive tests for each relevant metal sheet combination. For all subsequent spot welds, the thickness of the nugget can then be estimated nondestructively from the attenuation of the corresponding back wall echo.

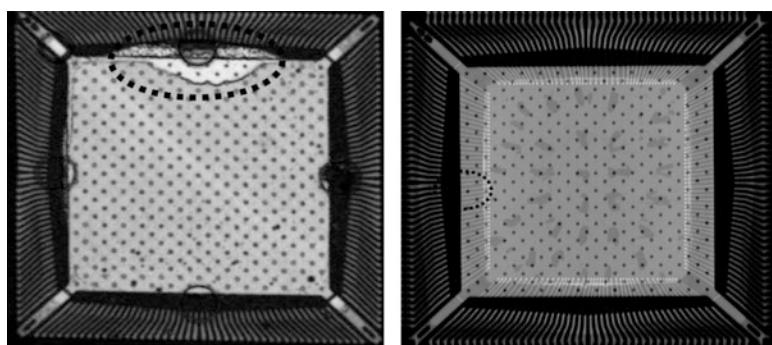
## Advanced 3D SAM Imaging

Various NDE methods are available to provide information about the presence, size, and orientation of flaws in an object under investigation. The NDE method to be selected depends on a variety of parameters, a few of them are: the properties of the

material to be tested (electrical conductivity, ferromagnetic properties, x-ray absorption coefficient, degree of heterogeneity, etc.), the type, position, orientation and size of the flaws expected, and the size of the components to be tested. Commonly several methods can be applied. Each of them has its own advantages and disadvantages and is able to image different object features.

A good example of the complementarity of different methods is given by (digital) X-ray radiography and scanning acoustic microscopy. The results of both methods are compared in Fig. 28 for the case of an electronic component containing a chip with a delamination. The images contain fairly identical information concerning a lot of object features, but several other features are visible in only one of the images. For example, the delamination, marked by the dashed black ellipse in the left picture of Fig. 28, is “seen” by the ultrasound clearly but is not at all visible in the X-ray radiography. This is well known and understandable: the large jump in the acoustic impedance produced by an air-filled crack or a delamination leads to a reflection of nearly all of the acoustic energy which can be detected by the transducer and displayed in the SAM image; on the other hand, the X-ray radiography is sensitive to changes in the absorption coefficient along the ray path – a value which does change significantly for a crack only if it is oriented parallel to the ray path. The radiography in turn shows features which are not present in the SAM image, especially the fine wires at the border of the chip connecting it to the frame (dashed black ellipse in the right picture of Fig. 28).

X-ray computed tomography (CT) provides additional 3D volume information based on the density distribution within the object and this data can be visualized by rendering software. The 3D capability allows a very easy and intuitive interpretation of the measurements. While several other NDE methods like SAM also possess 3D information in principle, this information is usually not of the same quality and accuracy as that of the X-ray CT. However, this situation can be significantly improved as shown in the following subsections (Barth et al. 2008).



**Fig. 28** Two images of a microchip containing a stress-induced delamination. On the left: Scanning acoustic microscopy at 80 MHz; on the right: X-ray radiography

### 3D Imaging in Scanning Acoustic Microscopy

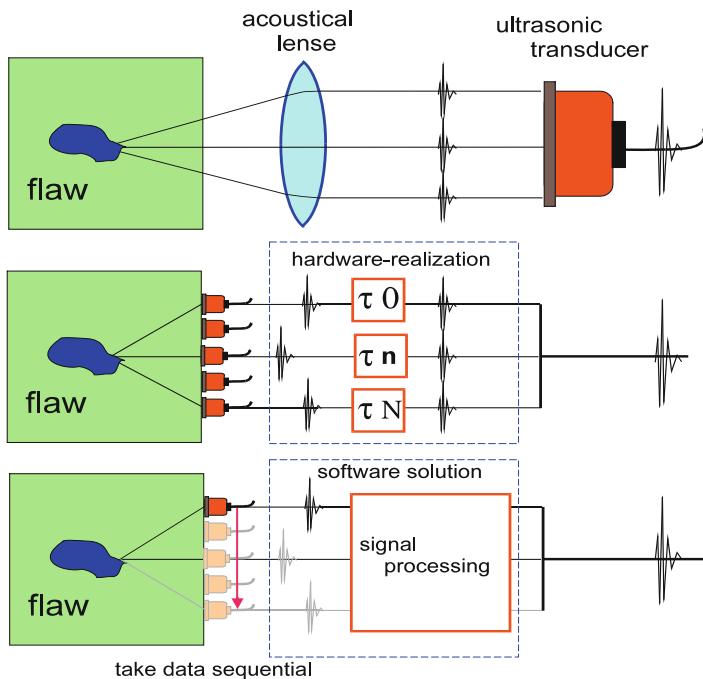
Besides several scan modes such as C-Scan (essentially representing a horizontal cross section of the object) and B-Scans (vertical cross section), and variants thereof, normally a scan mode collecting a 3D data set is also offered. However, there is usually no rendering software similar to the CT available because the acoustic data set does not directly represent volume properties which can be connected to the internal properties of the object. There are several reasons for that:

- (i) To get a reasonable resolution usually focused transducers are applied. Due to the limited focal depth (length of the focal area), a high lateral resolution is only achieved in a small depth range. The data at other depths do not accurately represent the volume information.
- (ii) If there are inclined cracks or delaminations in the volume, the ultrasound is reflected sideways. These echoes are not detected by the transducer and thus, the corresponding features are usually not represented in the data set. However, the echoes of inclined faces are detectable, if the transducer (acting in pulse-echo mode) is in a position laterally displaced from the scatterer. This means that the scattered echoes are represented but at wrong positions in the data set.

There exist other ultrasonic imaging concepts based on direct coupling of transducers to the object surface, which indeed have volume visualization capabilities, e.g., the phased array techniques and synthetic aperture focusing techniques (SAFT, see e.g., Langenberg et al. 1986; Mayer et al. 1990). It is the aim of this subsection to show that appropriate extended concepts are also applicable to acoustic microscopy. For many practical applications this would be a good alternative to X-ray CT avoiding all of the disadvantages connected with ionizing radiation as well as administrative and technical measures connected with radiation protection. Contrary to the X-ray CT there is a real chance to visualize delaminations and inclined cracks with small crack openings as well.

### General Focusing Techniques in Ultrasonic NDE

In acoustic microscopy a high axial and lateral resolution is necessary. For this purpose, the acoustic waves are focused to a given depth by an acoustical lens or by an appropriate shaping of the transducer surface. In both cases, the central rays are delayed compared to the outer rays. In the focal depth, they arrive at the same time (in phase) due to the longer travel path of the off-center rays. This is illustrated in Fig. 29 (top row). However, this approach is not very flexible. If the whole volume has to be scanned with high resolution, the 2D lateral scans have to be repeated several times with the focus at various depths. Also beam steering is not possible, so that defects are illuminated more or less from the incident plane leading to problems in detection of inclined crack surfaces.



**Fig. 29** Three of the possible methods to focus ultrasound into a sample. The focusing can be applied on both the sending and receiving side. Top: Traditional SAM focusing; Centre: Phased array focusing; Bottom: Synthetic aperture focusing

For direct coupling two other – more flexible – methods are available to focus the ultrasound to a given point in the specimen. In phased arrays, a 1D (line) or 2D (matrix) array is excited with time delays between the transducer elements provided by electronic hardware (Fig. 29, mid-row). Within the limits of wave diffraction, fast focusing to all depths and beam steering is possible by electronic manipulations. One essential limitation is that the resolution is within the range of the size of the transducer elements (Stepinski 2007). Therefore – as a rule – the element size is chosen in the order of (half of) the ultrasonic wavelength. Yet smaller elements (at the same frequency) do not make sense due to the diffraction limit.

Another way to reach the same goal is based on moving a small transducer along the surface over a *synthetic* aperture. All data are collected in the memory of a computer and a numerical delay and sum algorithm is applied afterwards (Fig. 29, bottom row). This can be done for all volume points (voxels) in the sample. Also in this case the (real) transducer aperture determines the lateral resolution in the images (together with the size of the synthetic aperture and the signal frequency).

In acoustic microscopy so far only the acoustic focusing by a lens or a curved transducer is commonly used whereas the focal spot is usually positioned near the depth of the object to be displayed. The procedures indicated in Fig. 29 (mid-row and bottom row) are only implemented for “low frequency” applications up to 20 MHz to the

best of the authors' knowledge. The reason might be that usually for these techniques, which yield high resolution by combining the signals of a large number of small-size transducer elements, the resolution is limited by the transducer element size (amongst others). SAM aims for a resolution significantly below  $100\text{ }\mu\text{m}$  and it is difficult to get a transducer element size of this order. The SAM tomography described in the following is based on the third case but without using a transducer with a small real aperture. Instead a "virtual transducer" consisting of the focal spot of a (larger) real transducer is used. This focus is placed at or near the sample surface as was already demonstrated in Fig. 2.

### Test Specimen and X-Ray CT Image

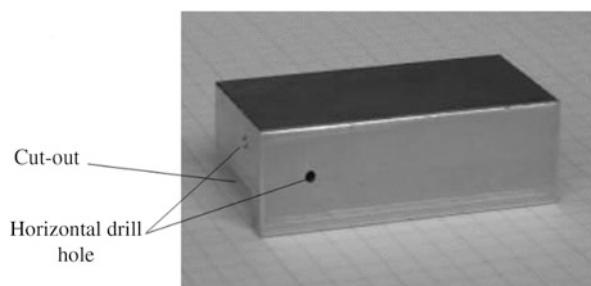
For a better comparison, a test specimen was machined out of a small block of aluminum which is appropriate for both X-ray CT and SAM tomography measurement (Fig. 30). The sample contains two flat bottom holes drilled in the horizontal direction and perpendicular to each other as well as a number of flat bottom holes (not visible in Fig. 30) with increasing inclination. One flat bottom hole was drilled from the back side perpendicular to the surface. The diameter of the horizontal holes was 1.5 mm and the diameter of the remaining holes was 1 mm in each case. The size of the sample is small enough to allow X-ray CT with a standard energy of 150 keV.

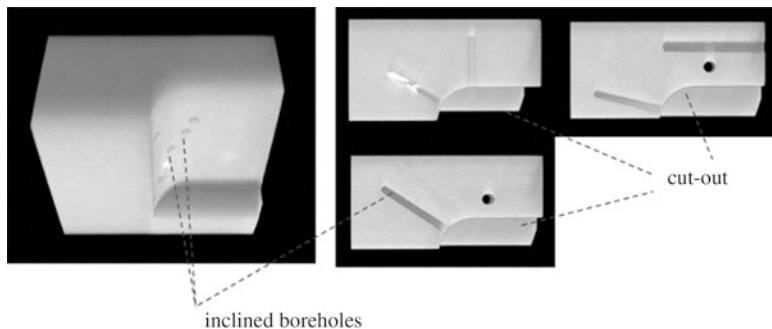
Figure 31 gives rendering and cross-sectional results of the CT. All object features are clearly visible. Of particular interest is a broken steel drill, which got stuck in one of the holes and is displayed bright white due to the much larger absorption coefficient of steel compared to aluminum. In the following, it is shown that SAM tomography is able to produce results comparable to X-ray CT images.

### Signal Preprocessing and Ultrasonic Reflection Tomography

A commercial SAM was used to collect the ultrasonic pulse-echo data. The instrument allows to store for all scan points (2D array) the complete time signal (third dimension) and to export this data for external processing. For the measurements, a transducer with a frequency of 30 MHz was focused to the surface. An area of  $30\text{ mm} \times 25\text{ mm}$  was scanned. The number of scan points was  $1000 \times 833$ . In Fig. 32 (top picture), a typical set of measured raw data for one scan line is given.

**Fig. 30** Aluminum test sample with dimensions of  $50\text{ mm} \times 25\text{ mm} \times 15\text{ mm}$ . The visible holes have a diameter of 1.5 mm and a depth of 15 mm in each case





**Fig. 31** Results of the X-ray CT for the aluminum test sample. On the left: 3D rendering of the object; on the right: three vertical cross sections chosen in such a way that the drill holes become visible. The drill holes starting from the cutout are inclined and have a diameter of 1 mm. The diameter of the horizontal drill holes is 1.5 mm

The horizontal lines in the upper part marked in red represent the surface echo. The scaling is chosen such that the weaker signals in the image are also visible; therefore the surface echo is overamplified.

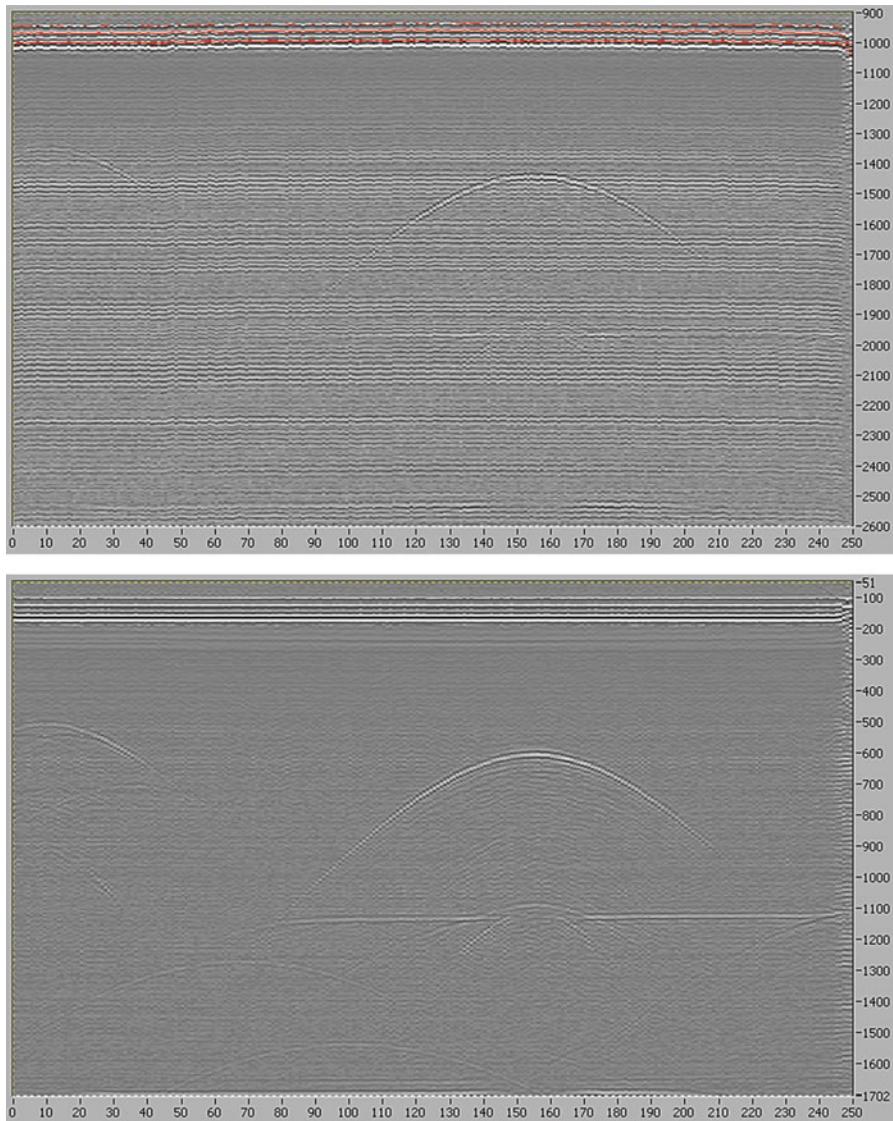
The horizontal lines in Fig. 32 (top picture) below the surface echo obviously do not correspond to structures in the specimen. Therefore they should better be removed prior to tomographic reconstruction. Figure 33 (top picture), giving the final reconstruction based on the raw data, confirms this assumption. The tomographic reconstruction is based on the algorithms described in Kak and Slaney (1988), Langenberg et al. (1986), and Mayer et al. (1990).

Several signal processing steps have been developed and implemented to improve the signal-to-noise ratio. Here they are shortly summarized:

- Noise reduction by filtering out frequencies without useful information.
- Removal of the jitter in the data acquisition.
- Removal of probe reverberation by subtraction of an appropriate reference signal.
- Retriggering to the surface echo.
- Removal of the surface echo reverberation.

All the listed signal processing steps turned out to be necessary and their correct sequence is essential. In Fig. 32 (bottom picture) and Fig. 33 (bottom picture), the improvement of the signal-to-noise-ratio and of the reconstruction result due to the preprocessing is obvious. The latter removes nearly all of the “clouds” which mask most of the relevant features. Especially the horizontal boreholes are completely invisible in the reconstruction based on raw data and are clearly visible in the reconstruction based on preprocessed data.

In Fig. 34, several scatterers in the SAM tomography are identified. The horizontal boreholes ( $d = 1.5 \text{ mm}$ ,  $0^\circ$ ) and inclined boreholes up to an inclination angle of  $45^\circ$  to the incident plane are clearly visible. Even the borehole with an inclination of  $60^\circ$  can still be identified. The side walls of the holes with  $75^\circ$  and  $90^\circ$  inclination

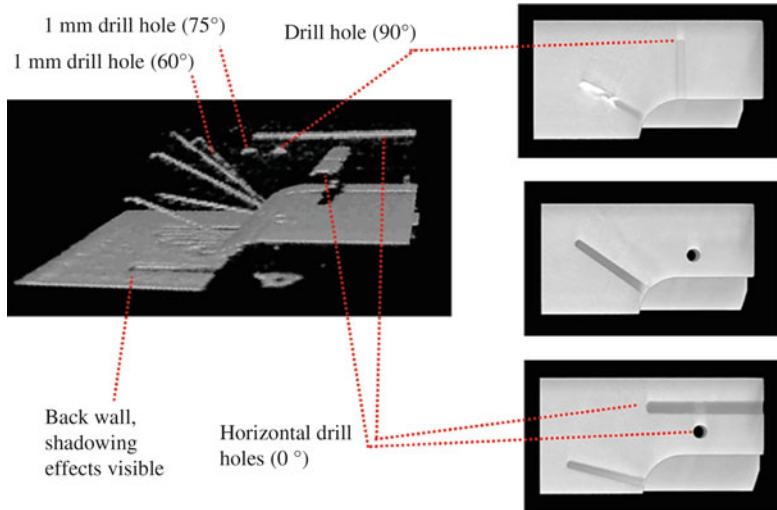
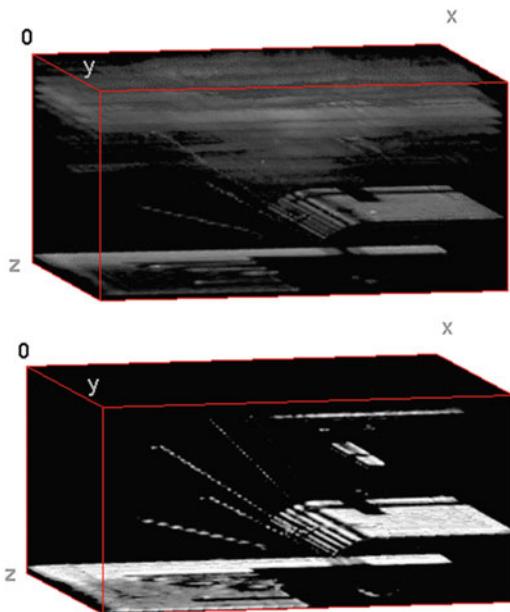


**Fig. 32** Raw data of a single scan line without preprocessing (top) and with preprocessed data (bottom). The scan direction contains 1000 scan points according to a total scan length of 30 mm. The data which correspond to the area above the surface are not displayed

are not visible. The sound field behind the focus has a finite beam opening angle and is obviously not wide enough to give echoes from such steep faces. However, the bottom of these holes is still visible and indicated.

Moreover, in Fig. 34 also shadows of the drill holes are present in the image of the back wall. At these surfaces specular reflection of the waves takes place. Thus, only

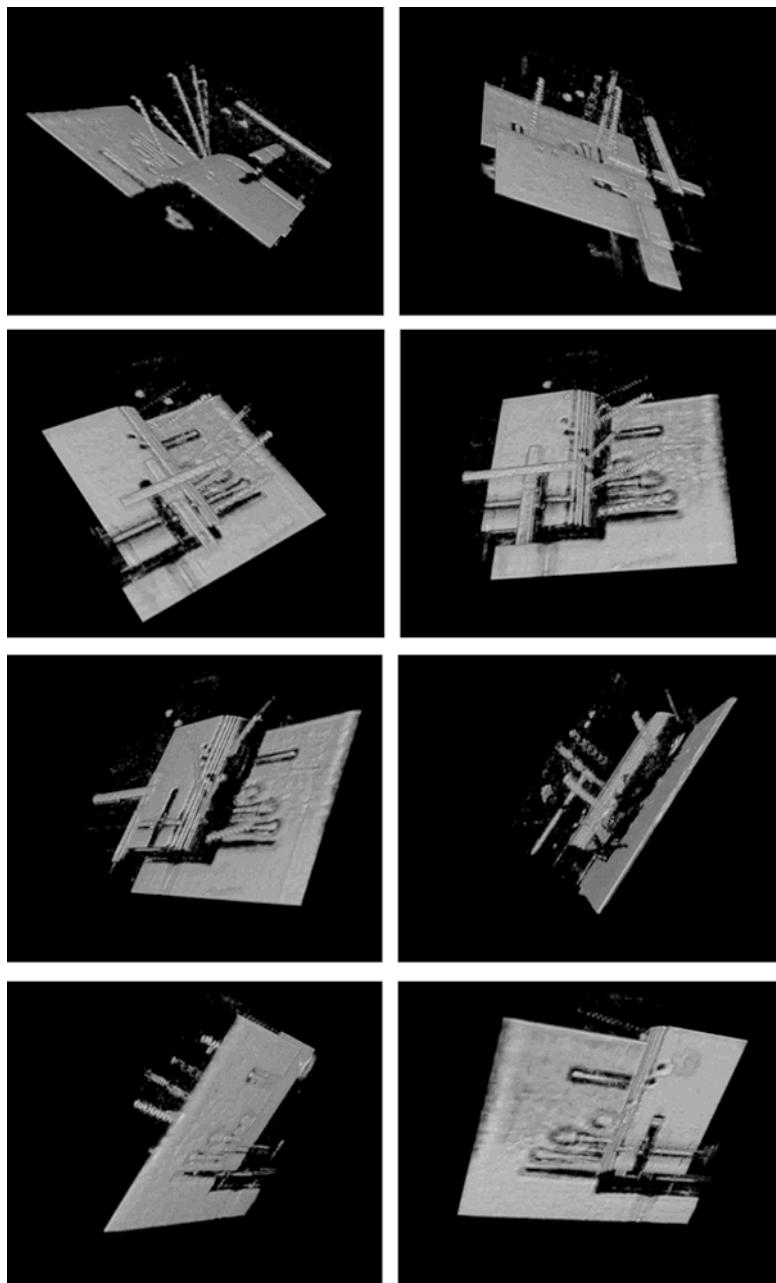
**Fig. 33** 3D Tomographic reconstruction based on raw data (top) and on preprocessed data (bottom)



**Fig. 34** Another rendering of the SAM tomography results (left hand side) compared with corresponding X-ray CT cross sections (images on the right)

the transducer position just above a given point of the horizontally orientated back wall contributes to that point. If there is another reflector in between, the back wall signal is blocked.

Figure 35 shows various snapshots of an animation sequence of the 3D SAM reflection tomography results. The pictures display the volume of the aluminum



**Fig. 35** Various snapshots of an animation sequence of the 3D SAM tomography showing the aluminum sample from different angles of view

sample shown in Fig. 30 from different angles of view. Once again the shadowing of the back wall due to the drill holes is clearly visible.

SAM tomography with a synthetic aperture seems to be a promising option to conventional C-Scan imaging if numerous defects at different depths are expected or if the depth of a single scatterer is not known a-priori. Moreover, since the beamforming capabilities of the SAFT algorithm are correlated with the near field of the synthetic aperture and this near field can be extended by enlargement of the synthetic aperture, a better focusing to very large depths is possible.

## Advanced SAM Imaging with Conical Annular Phased Arrays

Another solution to the deep focus challenge in SAM imaging is the well-known annular phased array. It consists of several concentric annular elements that can be separately controlled by a multichannel electronics (Fig. 36, top picture).

The annular array generates torus-shaped waves that naturally concentrate at the longitudinal axis. The behavior of the entire array can be controlled by pulse and receive delays (focal laws). With a fixed focal law the array acts like a probe with fixed focus distance. By changing the focal law, the focus distance and the focus shape can be changed electronically without mechanically moving the transducer up and down. However, due to the fact that the curvature of a conventional annular array is fixed to a predefined radius and thus, to a predefined focal depth, the flexibility of focus tuning is limited. For this reason, it is essential to combine the concept of an annular array with a conical form of the aperture as shown in the bottom picture of Fig. 36.

With a conical (or inverse axicon) annular array, a Bessel-like sound beam with a long and slim focus and significantly reduced geometrical spreading and diffraction

**Fig. 36** Annular phased array with 16 independent elements of the same area (top). The central element is omitted for a better sound field profile. The aperture is formed like a conus or an inverse axicon (bottom picture)

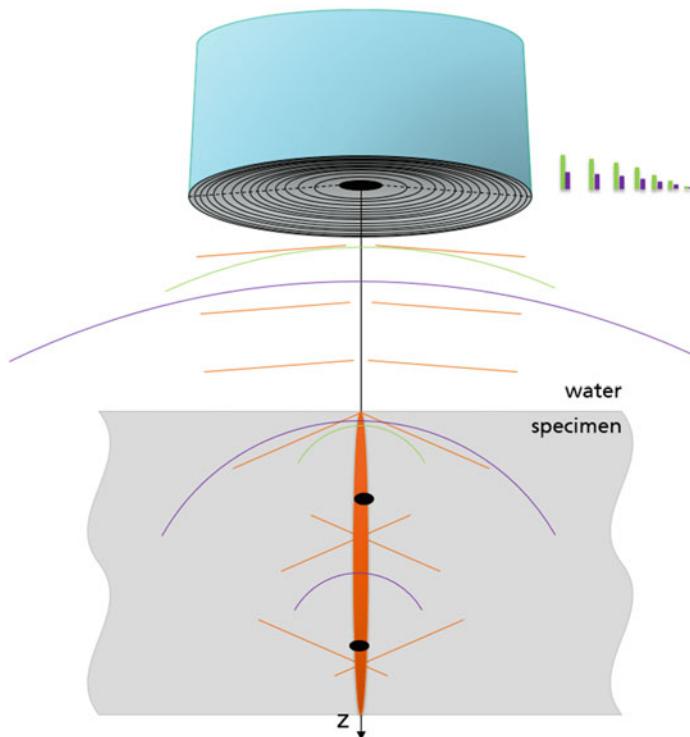


is generated (Fig. 37). The depth and length of the focal region is tunable by adaptive focal laws. On the receive side dynamic, i.e., time-of-flight or depth-dependent delays can be used in order to focus simultaneously to all depths where echoes may come from. As a result, a summed A-scan of a large depth range is obtained. Another advantage of the Bessel beam is the fact that the waves recreate the focus behind small obstacles so that shadowing effects are reduced.

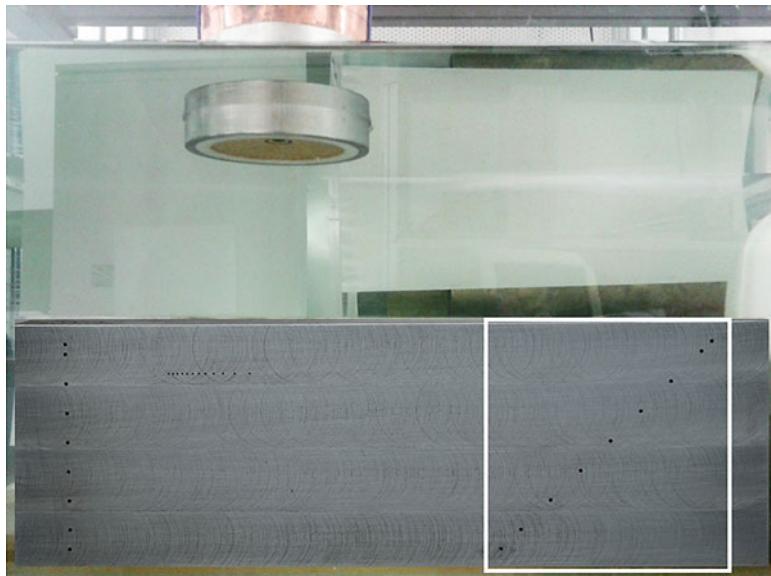
In order to test the principle of a conical annular array, a first prototype with a center frequency of 9 MHz and a diameter of 70 mm was designed and manufactured (Fig. 36, top). Its target application was the volume scanning of forged aircraft engine components with a thickness up to 76 mm (3 inches) and with flaws (flat bottom holes) down to 0.2 mm.

The probe and the underlying focal law procedure was tested in a SAM environment using a steel block with a height of 125 mm and with artificially introduced side drill holes of 2.5 mm diameter located in various depths (Fig. 38).

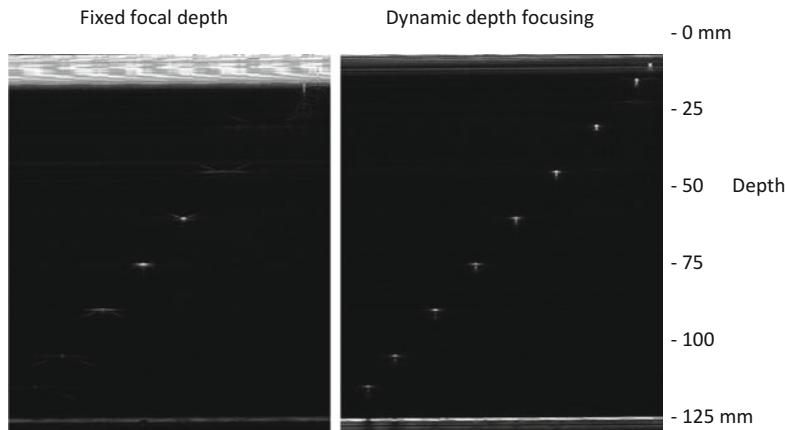
In Fig. 39, the results of two line scans across the region of interest as indicated by the white rectangle in Fig. 38 are shown. The first scan (picture on the left in Fig. 39)



**Fig. 37** Schematic principle of an annular array with conical aperture producing a long and slim sound beam with reduced geometrical spreading and diffraction (orange lines). With dynamic depth focusing the contributions from flaws in various depths (indicated by the green and purple circular arcs) can be summed up constructively at once, i.e., in one and the same A-Scan

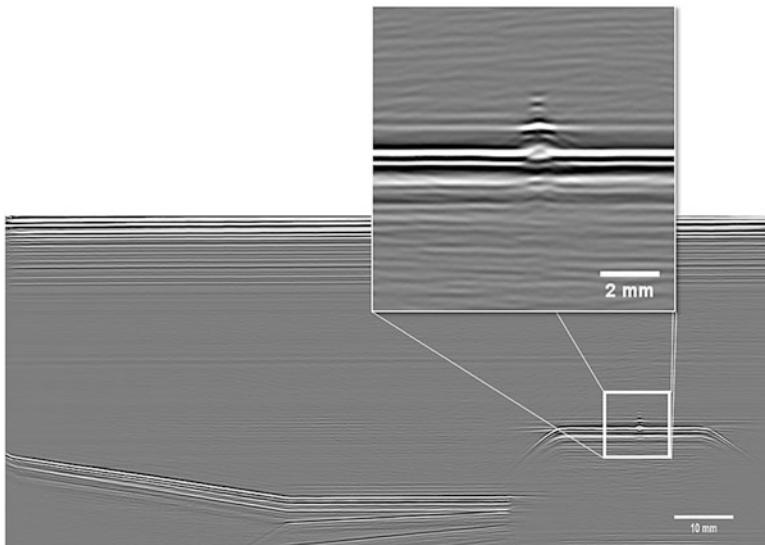


**Fig. 38** SAM measurement with dynamic depth focusing using the conical annular array. The scanned region of interest of the steel block with nine side drill holes in various depths between 10 and 120 mm is indicated by the white rectangle



**Fig. 39** SAM scan of the marked area with a fixed focal depth (on the left) and with dynamic depth focusing of the conical annular array (on the right)

was performed with a fixed focal depth as in conventional SAM. Only a limited number of the drill holes are visible and only one scatterer (the one at the focal depth) is imaged properly as a compact spot. In the second scan of the same area, dynamic depth focusing was used for all relevant depths (picture on the right). In this case, all



**Fig. 40** Identification of a 0.2 mm flat bottom hole located very close to a geometrical step of a forged aircraft engine component in a depth of 35 mm using dynamic depth focusing with a conical annular array at 9 MHz center frequency

nine drill holes are clearly visible and they are imaged with approximately the same size and brightness.

From Fig. 39, it can also be found that with dynamic depth focusing (DDF), the dead zone in the region near the fluid-solid interface is significantly reduced so that also near-interface defects can be identified. Moreover, the back wall of the specimen is imaged more clearly and with less shadowing effects.

In a further scan of a forged aircraft engine component, it was proven that with dynamic depth focusing flat bottom holes of 0.2 mm diameter can be found even if these scatterers are located at larger depth and very close to geometrical steps (Fig. 40).

The results obtained with the prototype of a conical annular array revealed that the DDF concept is physically sound and offers a significantly better performance at small and large depths than any other conventional probe so far. The technological challenge now is to scale the transducer to higher frequencies between 30 and 70 MHz so that a larger scope of SAM applications can be addressed.

---

## Summary

The scanning acoustic microscope (SAM) represents a multipurpose high-precision NDE tool for detailed analysis of material structures and components in a laboratory environment. Its main advantages are a large variety of exchangeable ultrasonic transducers and thus, a huge field of different applications, a very precise scanning

device, and an immersion approach with stable and reproducible coupling conditions. It therefore represents a reference UT system for evaluation and improvement of field-testing devices, for expert opinions but also for baseline investigations.

A SAM can further be seen as a cross-sectional ultrasonic NDE technology affecting a wide range of UT aspects like sound field simulation, wave-defect interaction, transducer design, ultrasonic electronics, scanner technology, signal processing, and imaging techniques. Any ongoing progress in one of these fields will automatically lead to an improvement and extension of SAM technology offering new perspectives for future applications.

Beside conventional A-, B- and C-Scan pictures, the SAM offers the possibility for advanced imaging techniques based on synthetic apertures (SAFT) and conical annular arrays (DDF). It is expected that in the near future other phased array technologies based on linear and matrix arrays together with corresponding reconstruction algorithms will be integrated so that new sophisticated imaging capabilities for SAMs will be provided.

**Acknowledgments** The authors would like to thank Peter Krueger and Sven Fischer of IKTS (former IZFP-D) for the X-ray CT of the specimen in section “[Advanced 3D SAM Imaging](#)” and for helpful discussions.

---

## References

- Barth M, Schubert F, Koehler B (2008) Where X-ray imaging fails – Delamination, crack, and micro-pore detection using ultrasonic reflection tomography in a scanning acoustic microscope, IEEE nuclear science symposium and medical imaging conference record 2008. In: Workshop on X-Ray micro imaging of materials, devices, and organisms, Dresden, Germany, on CD, pp 577–581
- Briggs GAD (ed) (1992) Acoustic microscopy. Oxford University Press, Oxford
- Briggs GAD (ed) (1995) Advances in acoustic microscopy, vol 1. Plenum Press, New York and London
- Fellinger P, Marklein R, Langenberg K-J, Klaholz S (1995) Numerical modeling of elastic wave propagation and scattering with EFIT – Elastodynamic finite integration technique. Wave Motion 21:47–66
- Gilmore RS (1999) Industrial ultrasonic imaging/microscopy. In: Thurston RN, Pierce AD, Papadakis EP (eds) Physical acoustics. Academic Press, New York, pp 275–346
- Kak A C, Slaney M (eds) (1988) Chapter 8: reflection tomography. In: Principles of computerized tomographic imaging IEEE Press, New York
- Langenberg K-J, Berger M, Kreutter T, Mayer K, Schmitz V (1986) Synthetic aperture focusing technique signal processing. NDT Int 19:177–189
- Lemons RA, Quate CF (1979) Acoustic microscopy. In: Mason WP, Thurston RN (eds) Physical acoustics. Academic Press, London, pp 1–92
- Maev RG, Denisov AA, Paille JM, Shakarji CM, Lawford BB (2004) Spot weld analysis with 2D ultrasonic arrays. J Res Natl Inst Stand Technol 109(2):233–244
- Maev R G, Denisov A A, Erlewein J., Roemmer H (2011) Advanced ultrasonic imaging for automotive spot weld quality testing, In: Conference proceedings of 5th Pan American conference for NDT
- Mayer K, Marklein R, Langenberg K-J, Kreutter T (1990) Three-dimensional imaging system based on Fourier transform synthetic aperture focusing technique. Ultrasonics 28:241–255

- Schubert F, Peiffer A, Koehler B, Sanderson T (1998) The elastodynamic finite integration technique for waves in cylindrical geometries. *J Acoust Soc Am* 104(5):2604–2614
- Schubert F, Koehler B, Peiffer A (2001) Time domain modeling of axisymmetric wave propagation in isotropic elastic media with CEFIT – cylindrical elastodynamic finite integration technique. *J Comp Acoust* 9(3):1127–1146
- Schubert F, Koehler B, Zinin P (2005) Numerical time-domain simulation of wave propagation and scattering in acoustic microscopy for subsurface defect characterisation. In: Geer RE (ed) Testing, reliability, and application of micro- and nano-material systems III: proceedings of the third conference on testing, reliability and application of micro- and nano-material systems, San Diego, California, SPIE proc. series 5766. Society of Photo-Optical Instrumentation Engineers -SPIE, Bellingham/Wash, pp 106–117
- Schubert F, Hipp R, Gommlich A (2014) Determination of diameter and thickness of weld nuggets in resistance spot weldings by high-frequency ultrasound inspection. In: Proceedings of the 11th European conference on nondestructive testing (ECNDT), Prague, on CD
- Stepinski T (2007) An implementation of synthetic aperture focusing technique in frequency domain. *IEEE Trans UFFC* 54:1399–1408
- Zhu J, Popovics J, Schubert F (2004) Leaky Rayleigh and Scholte waves at the fluid-solid interface subjected to transient point loading. *J Acoust Soc Am* 116(4):2101–2110
- Zinin PV (2001) Quantitative acoustic microscopy of solids. In: Levy M, Bass H, Stern R, Keppens V (eds) Handbook of elastic properties of solids, liquids, and gases. Vol I: dynamic methods for measuring the elastic properties of solids. Academic Press, New York, pp 187–226
- Zinin PV, Weise W (2003) Theory and applications of acoustic microscopy. In: Kundu T (ed) Ultrasonic nondestructive evaluation: engineering and biological material characterization. CRC Press, Boca Raton



# Acoustic Emission

4

Eleni Tsangouri and Dimitrios G. Aggelis

## Contents

Introduction .....	116
AE Principles .....	116
Brief History of AE Use in Monitoring Concrete Structures .....	118
AE Analysis Tools .....	119
Hits Distribution .....	119
Source Localization .....	120
Wave Features .....	122
Case Study: AE Monitoring of Damage and Repair of Self-Healing Concrete .....	123
Material Design .....	125
Healing Efficiency Assessment .....	125
Test Setup .....	127
AE Setup .....	127
Results .....	127
Discussion and Future Perspectives .....	133
Summary .....	133
Problem of Signal Complexity .....	135
Synergy with Other NDE Techniques .....	136
Future Perspectives .....	136
Advantages and Limitations of AE .....	137
References .....	138

## Abstract

Acoustic Emission (AE) is an NDE technique that measures the elastic energy released in the form of acoustic waves in materials that undergo irreversible structure changes (e.g., plastic deformation, cracking, rupture). The emitted acoustic waves are captured by sensors attached to the material's surface. The

E. Tsangouri · D. G. Aggelis (✉)

Department of Mechanics of Materials and Constructions (MeMC), Vrije Universiteit Brussel (VUB), Brussels, Belgium

e-mail: [eleni.tsangouri@vub.be](mailto:eleni.tsangouri@vub.be); [etsangou@vub.ac.be](mailto:etsangou@vub.ac.be); [dimitrios.aggelis@vub.be](mailto:dimitrios.aggelis@vub.be); [daggelis@vub.be](mailto:daggelis@vub.be)

technique considers the wave arrival time and the wave propagation speed to accurately locate in space and in time the wave source and the waveform shape to characterize the origin of nature (damage mode, crack orientation, etc.). AE monitoring systems have the ability to effectively inspect large and complex structures, are cost-effective and perform in the long-term. For these reasons, AE is the most commonly implemented NDE method in civil engineering and especially in concrete structures inspection. For several decades, AE is employed in numerous studies to fully assess and characterize cracks nucleation and propagation, damage mode and source on concrete. As concrete design and technology evolves, AE application is challenged. A case study application of AE highlights the recently developed smart autonomously healed concrete structures, the “concrete of the future.” It is shown that AE aims to develop a sensing tool; a sensor array is sufficient to detect, locate (and trigger if needed) the healing activation processes, quantify damage before and after repair. An overview of studies performed in the recent years shows the performance of AE on healing systems carrying agent that is encapsulated or embedded into advanced vascular networks.

---

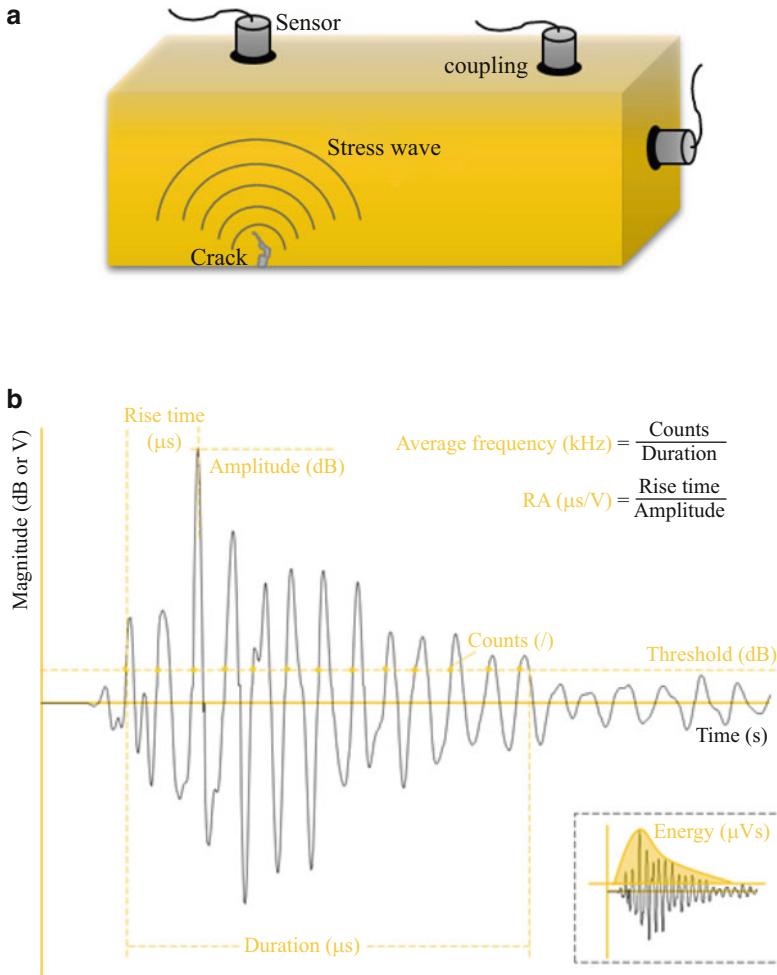
## Introduction

### AE Principles

As stress is applied to a material, strain is induced due to atomic-scale movement. When deformation leads to cracking, transient elastic waves are radiated from the source under stress and spread in all directions. Acoustic Emission (AE) technique monitors this physical phenomenon by detecting these waves that travelled through the material and reached its outer surface.

AE was used as structural integrity assessment method since antiquity: as early as 6500 BC potters listened to audible sounds emitted during earthenware casting at the end of the cooling process (Shull 2016). In modern days, AE was introduced as emerging NDE method with the PhD dissertation of Joseph Kaiser (Kaiser 1953) who determined the phenomenon of AE irreversibility in materials under stress.

Synchronous AE software consists of an array of sensors, signal preamplifiers, noise filters, data acquisition, display, and storage equipment. The sensors are made of piezoelectric elements (Fig. 1a). This highly sensitive sensor is mounted on the material surface through a coupling agent, captures the transmitted elastic waves, and converts them into electric voltage signals (Fig. 1b). Based on the performance of the sensors along the frequency spectrum, they are grouped in two classes: resonant sensors that are highly sensitive to narrow frequency bands and broad or wideband sensors that perform equally well (but with lower peak sensitivity compared to resonant ones) in a wide range of frequencies. The sensor is connected to a preamplifier that boosts the signal voltage (Hsu-Nielsen source calibration method is used to ensure that the sensors are in good acoustic contact with the material surface).



**Fig. 1** (a) Typical AE setup; (b) AE signal hit features

This standard method consists of pencil-lead breakages that generate a burst, fast in rise time, intense and remarkably reproducible signal similar to natural AE source.).

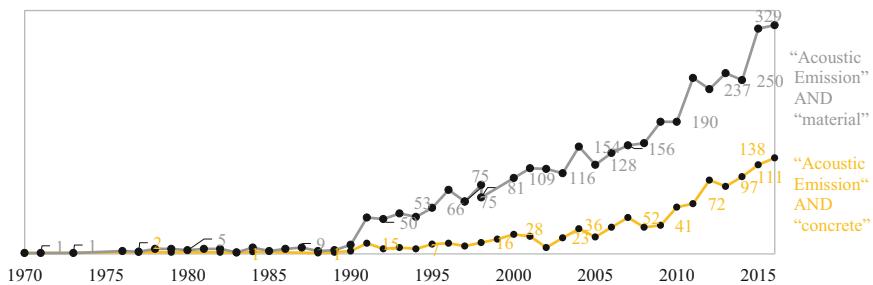
The most popular AE monitoring softwares at present use the hit-driven architecture: a “hit” is used to describe the signal captured by a sensor (schematically shown in Fig. 1b). A hit is defined as an acoustic signal with amplitude greater than the preset voltage threshold (Fig. 1b), set to eliminate noise effects. Namely, in the case that an elastic wave (AE event) is emitted and captured by a number of sensors, the AE acquisition system should detect the respective number of hits, one from each sensor. The signal shape and hit statistical information are collected and stored. A series of AE signal features are extracted (schematically shown in Fig. 1b).

## Brief History of AE Use in Monitoring Concrete Structures

Since the early 1960s when concrete was established as the dominant construction material, AE was applied to accurately monitor concrete performance and provide accurate and integrated insight into the material (Dunegan et al. 1968). AE appeared to be the method that fully characterizes the material's response at different stages of concrete production: AE has been used to evaluate concrete curing processes and along its service life detecting the moment that damage emerges (Ohtsu 1987).

The first instrumentation introduced by Pollock in 1969 is designed to handle stress waves emitted by materials undergoing deformation (Pollock 1969). The method, currently named 'stress wave emission,' analyzes the waveform and frequency spectrum to assess the presence and nature of defects. Sporadic studies are documented going back to 1970s, when AE started attracting interest from the concrete industry. The ASTM standard in 1977 discussed the output of early studies that measure stress waves emitted as plain Portland cement concrete is subjected to external loads (Nielsen and Griffin 1977). In the 1980s, the technique gained popularity with the development of advanced analog-to-digital (A/D) signal converters capable of performing at high sampling rates and storing large number of signals (Shull 2016). Early studies used AE to detect microcracks in the brittle cement matrix that control the growth of macrocracks (Evans et al. 1976). Since then, AE analysis was combined with fracture mechanics in order to identify the heterogeneous nature of concrete and its complex failure (Rossi et al. 1990; Maji and Shah 1988; Otsuka and Date 2000). A series of studies applied AE location analyses to detect microcracks (e.g., aggregate-matrix detachment (Rossi et al. 1989)) and progressive damage in different material phases (aggregate-matrix interphase, matrix phase) (Landis 1999).

Up to 30% of the total number of research studies concern concrete material and structures. As concrete structures design evolved in time, AE was steadily applied to assess the structural performance of concrete. The rebars-concrete bond deterioration, the weak point of traditional steel rebars reinforced concrete, was determined by measuring accumulating AE activity in rebars pull-out tests (Kobayashi et al. 1980; Grosse et al. 1997). The durability of ordinary and prestressed steel rebars reinforcement was assessed by measuring the magnitude of acoustic emission under cyclic incremental loading (Shield 1997). The damage patterns (multiple micro-cracking due to fibers bridging that enhances material fracture toughness) on short fiber-reinforced concrete are accurately visualized based on AE source localization (Mobasher et al. 1990; Shah and Choi 1999). Eventually, AE appeared to be an efficient tool that determines the contribution and performance of fiber reinforcement used in the form of rebars (Mirmiran and Philip 2000) or external sheets/plates (Carpinteri et al. 2007). This brief retrospection on AE history leads to the early 2000 when, according to the graph in Fig. 2, a breakthrough in AE applications occurred. At that moment, more sophisticated AE analysis tools were developed that permit effective monitoring of damage even in complex concrete composite structures or elaborate external loadings. The main AE analysis tools are analytically discussed in the next paragraph.



**Fig. 2** The scientific interest in acoustic emission monitoring of materials performance progressively increases since 1990

## AE Analysis Tools

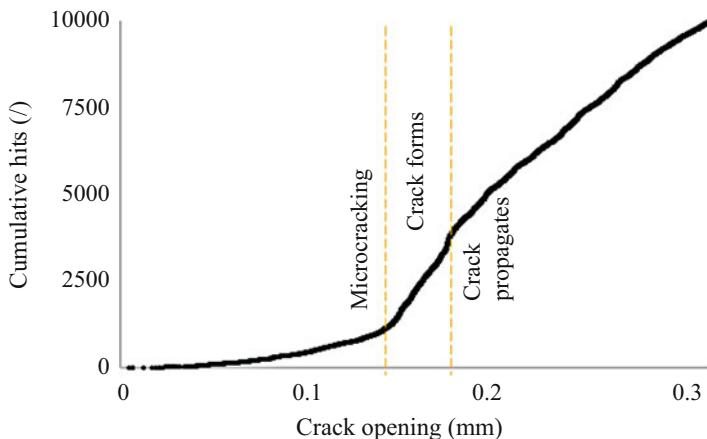
Three main AE data analysis approaches exist based on: (i) hits distribution, (ii) source localization, and (iii) wave features. The majority of AE studies consider one or a combination of these tools to assess material/structural integrity. The AE analysis tools are presented at the following paragraphs in details.

### Hits Distribution

The accumulation of AE hits occurs due to material internal irreversible changes: micro/macro-crack formation or propagation, particles friction, components local or global failure, interphase debonding, detachment, etc. The cumulative hits distribution and the hits rate measurement can be indicative of fracture. More analytically, the slope of the cumulative hits curve increases as the damage transits from micro to macro level. The transition point is referred as the ‘curve knee’ and defines the onset of numerous simultaneous damage phenomena that lead to unstable fracture propagation and final failure.

In concrete under stress, change in hits evolution can be instant (jump in value, i.e., interphase debonding) or progressive (hits release rate ranges in time, i.e., crack propagation). For instance, the cumulative hits distribution is given in Fig. 3 in the case of plain concrete that progressively cracks under bending. It is shown that hits evolve in three stages: (a) as concrete linearly deforms, negligible AE hits are released; (b) as microcracking occurs and fracture zone forms, hits accumulate and their population increases significantly; (c) as crack forms and propagates hits are continuously released, but at this stage the hits rate is lower.

Under incremental load, hits distribution is particular: in healthy state or in stable crack propagation mode, materials release no hits at load levels lower than the final load of the previous test cycle (Tsangouri 2015). The phenomenon, named the Kaiser effect demonstrates the irreversible nature of damage and acoustic emission. Plain concrete, rocks, and other brittle materials exhibit a Kaiser effect since they remain



**Fig. 3** Cumulative hits distribution on concrete beam under bending

relatively undamaged until the final strength of the sample whereby instant macro-damage occurs. On the contrary, materials with time-dependent mechanical behavior (i.e., polymer composites) or concrete carrying complex reinforcement (fibers or rebars) develop AE activity earlier than the previously exceeded stress levels (Aggelis et al. 2013). This reverse Kaiser phenomenon is called the Felicity effect.

Kaiser/Felicity effects are indicative of structural integrity (Rilem TC 212-ACD 2010). Rilem TC-212 ACD recommendation introduced Felicity ratio that quantifies the damage intensity at each load cycle (i):

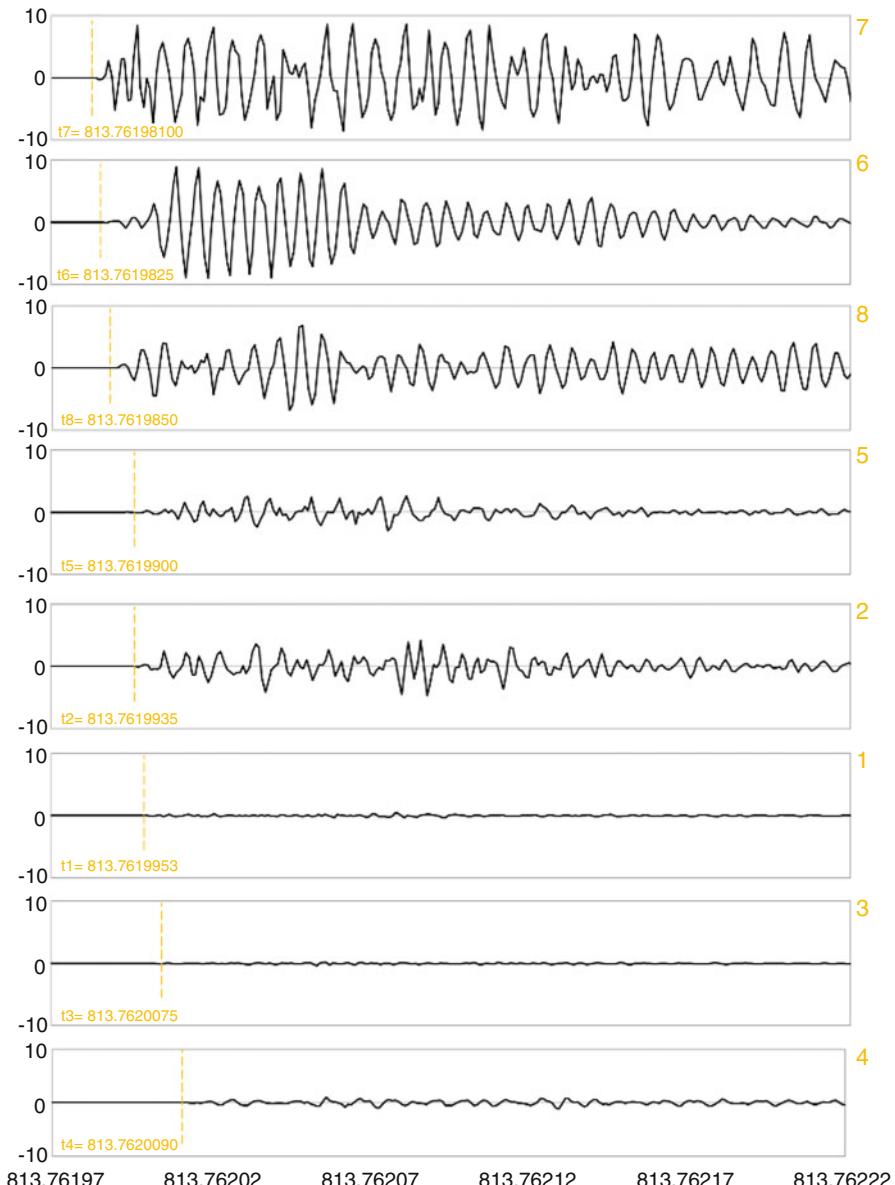
$$\text{Felicity ratio} = \frac{\text{Load at hits onset [Cycle}(i+1)\text{]}}{\text{Ultimate load [Cycle}(i)\text{]}}$$

$$= \begin{cases} \sim 1 & \text{Kaiser effect,sound and undamaged} \\ < 1 & \text{Felicity effect,unstable damage evolution} \end{cases}$$

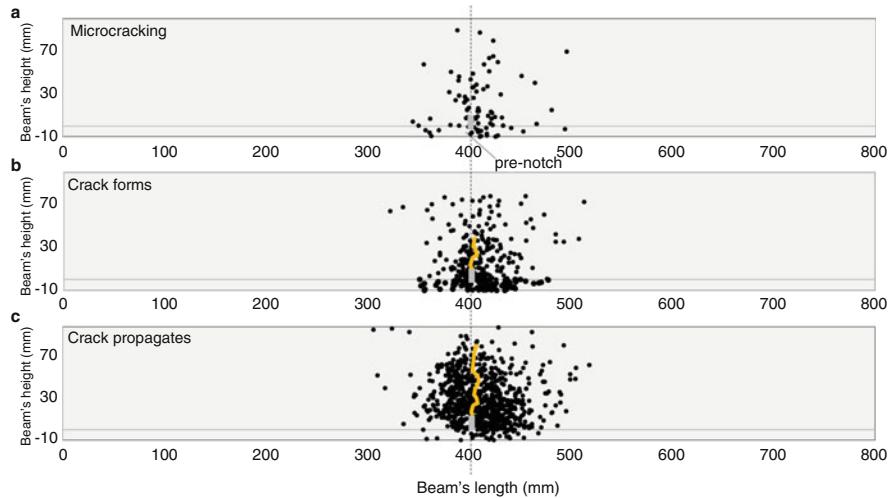
## Source Localization

Accurate localization of AE sources is a significant step toward damage evaluation. Source localization is an inverse problem that concerns the material wave propagation velocity (e.g., 4000 m/s for normal strength plain concrete) and the position of a sensors network and is processed by triangulation methods. Stress waves excited from a source (AE event) travel through the material and hit multiple sensors on its surface. The time delays between arrival times among sensors are used to detect source position. Considering that the unknown variables are the emission time and the source coordinates, at least one, two, three, and four sensors should be applied to perform zonal (events originated from a zone surrounding the sensor), linear (events

located along the one-dimensional sensor array), planar, and three-dimensional source localization, respectively. In case the signal is captured by more than the minimum sensors required, optimization methods are called to enhance the source localization accuracy. In Fig. 4, the signal hits captured by an array of eight AE sensors are presented and each arrival time is highlighted.



**Fig. 4** Waveforms captured by an array of eight AE sensors sorted by arrival times



**Fig. 5** AE events projected on the X–Y plane for the case of concrete beam tested under bending (Aggelis 2011)

In concrete studies, sensors are applied surrounding the zone prone to damage. AE can accurately locate crack nucleation and propagation in a broad range of applications: from long-span bridges to lab-scale beam bending tests. In Fig. 5, the AE events localized using a 3D localization algorithm with hits from 8 AE sensors are projected in a plane view. In agreement with concrete fracture mechanics models, a wide fracture process zone of microcracks is built up surrounding the macrocrack. At early loading stage, only a few events are localized, correlated to microcracking that builds the fracture process zone (Fig. 5a). The density of AE events increases significantly as a unique crack in the middle section (at 400 mm along the beam's length) forms and propagates (Fig. 5b, c).

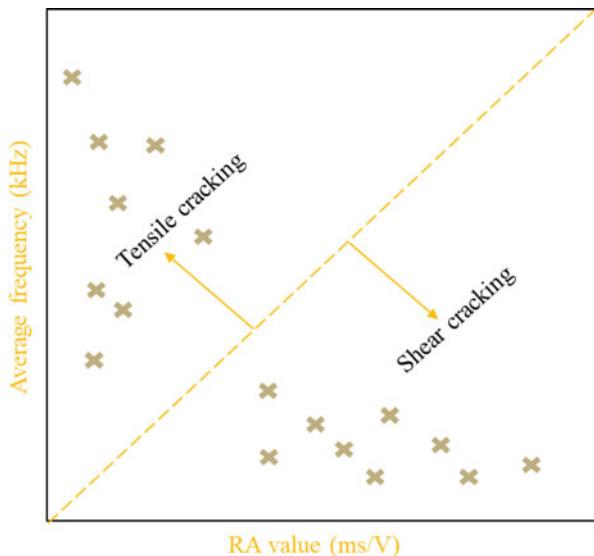
Accurate damage localization remains a challenging task since the commonly applied triangulation methods consider constant wave propagation velocity. However, in service and under stress, due to progressively developed defects the wave propagation path becomes complex and wave velocity degrades.

## Wave Features

The analysis of the relevant signal features (i.e., energy, duration, frequencies, amplitude, rise time, counts, etc.) provides crucial information for damage characterization. Representatively, the fracture energy released when plain concrete cracks is quantified by AE energy (Landis 1999). Counts, duration, and amplitude help to characterize the signal shape and magnitude.

Monitoring the AE features evolution can point out the dominant damage phenomena (e.g., energy rise defines the transition from micro to macro level damage) or

**Fig. 6** Protocol of Average frequency-RA value based on Rilem TC 212-ACD recommendation (Rilem TC 212-ACD 2010)



the damage mode (e.g., the rise time of signals emitted under shear damage mode is greater than the rise time of signals emitted under tensile damage mode) (Grosse et al. 1997; Tsangouri 2015).

Rilem TC 212-ACD recommendation establishes a test method based on AE features analysis for classification of active cracks in concrete into tensile and other-type cracks including shear cracks, see Fig. 6 (Rilem TC 212-ACD 2010). As illustrated in the graph, the RA value and the average frequency distribution is selected and considered to characterize damage source in concrete.

### Case Study: AE Monitoring of Damage and Repair of Self-Healing Concrete

The previous sections of this chapter aimed to introduce the AE theory, principles and analysis tools. In the following section, a case study is examined in which AE is the experimental technique that assesses damage and autonomous repair of a recently designed and developed innovative type of concrete. The material concept is briefly given and the most promising results of AE analyses are selected and discussed. The study aims to highlight the dominant contribution of the AE technique without getting into details regarding the self-healing concrete performance. Table 1 provides an overview of scientific articles focused on self-healing concrete mechanical performance in which AE is applied as monitoring tool. (The reader interested in this topic can find further information in recent Master (Thierens and De Valck 2015, Lelon 2017) and PhD dissertations (Tsangouri 2015; Van Tittelboom 2012).)

**Table 1** Studies overview focused on self-healing concrete mechanical performance at which AE is applied as monitoring tool

Author and Publication year	Test scale	Crack pattern	AE analysis tools and output highlights
(Minnebo et al. 2017)	Real	Multiple cracks	AE cumulative hits distribution assesses crack closure and reopening after healing; Source localization detects new cracks formed and cracks closure after healing
(Feiteira et al. 2017)	Small	Unique crack	AE cumulative hits distribution indicates agent fracture as cracks reopen after healing; AE features (i.e., energy) highlight agent-concrete interphase debonding
(Tsangouri et al. 2016)	Small	Unique crack	Source localization applied to assess AE accuracy in presence of cracks and accuracy restoration after healing
(De Belie et al. 2015)	Real	Multiple cracks	AE hits energy analysis detects healing activation in large-scale concrete beams
(Karaïskos et al. 2016)	Real	Multiple cracks	AE hits energy analysis detects healing activation in large-scale concrete beams
(Grosse and Malm 2016)	Small	Unique crack	Source localization detects capsule breakage; AE energy distribution assesses the fracture process on concrete
(Malm and Grosse 2016)	Small	Unique crack	AE energy analysis distinguishes the fracture processes of polymer healing agent from agent debonding and concrete cracking
(Van Tittelboom et al. 2016)	Real	Multiple cracks	AE hits energy analysis detects healing activation in large-scale concrete beams
(Van Tittelboom et al. 2015)	Small	Multiple cracks	AE hits energy analysis detects healing activation in concrete where multiple cracks form and interact after healing
(Tsangouri et al. 2013a)	Small	Unique crack	AE hits features (i.e., energy, duration) analysis localizes in space and in time capsule breakage events; Source localization applied to detect the area where healing agent is released
(Van Tittelboom et al. 2012)	Small	Unique crack	AE hits energy analysis detects glass capsules rupture as crack propagates
(Shiotani et al. 2009)	Real	Multiple cracks	AE Tomography localizes damaged and manually repaired zones in deteriorated dam structure
(Seo and Kim 2008)	Small	Multiple cracks	Cumulative AE energy and count applied to identify fatigue damage and healing in asphalt concrete and to evaluate the impact of rest period on damage complexity
(Granger et al. 2007)	Small	Unique crack	AE energy measurements on ultra-high-performance concrete characterizes the response to damage of newly formed crystals precipitating in the cracks after healing

## Material Design

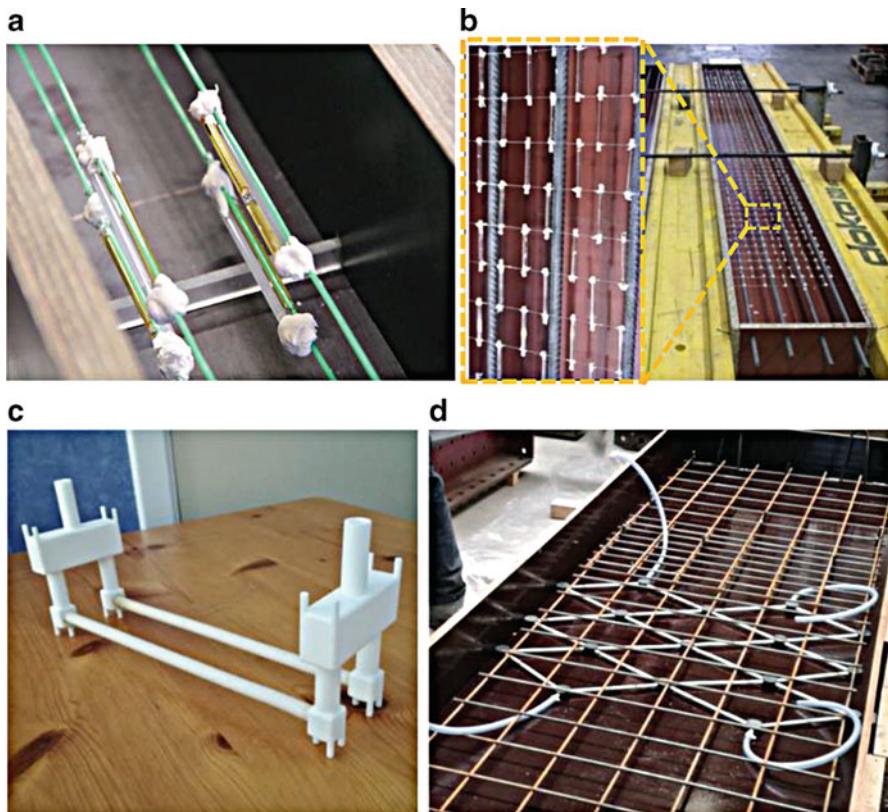
Self-healing materials form a new materials class (concrete, polymers, metals, ceramics, coatings) developed over the recent decade (Ghosh 2009). Self-healing concrete is a smart autonomous material: it has the ability to react to cracking by triggering embedded repair mechanisms. In detail, a healing agent is encapsulated and released only when damage occurs aiming to fill the crack void and restore mechanical integrity. Epoxy resin, cyanoacrylate, or alkali-silica solutions are the most efficient agents due to high rheology (low viscosity that permits cracks filling) and good adhesion to concrete cracked surfaces. Dry introduced a preliminary design of self-healing concrete in 1996 (Dry and McMillan 1996) and over the next years two encapsulation approaches were established:

**Short tubular capsules enclose the healing agent.** Hollow capsules are filled with agent embedded into concrete during mixing (Fig. 7a, b). The capsules are brittle enough to break as crack forms. The optimal capsules design is still under investigation (Minnebo et al. 2017). Glass, ceramic, and cementitious-based capsules are the most promising solutions due to good adhesive properties and brittle response to damage (Minnebo et al. 2017). The self-healing concept is characterized as passive since there is no need for human intervention in order to trigger the healing process.

**Continuous supply two-dimensional piping system.** Tubular pipes are placed into concrete at mixing and remain empty (Dry and McMillan 1996). The pipe breaks as crack forms and an internal/external reservoir system is triggered to deliver healing agent at the cracked pipe section (Fig. 7c). A vacuum pressure system ensures continuous agent supply in the case of repeatable cracks. The self-healing piping systems developed to date are considered active since human intervention is required in order to trigger agent delivery (Minnebo et al. 2017). Research studies aim to develop supply systems with enclosed sensing components that detect damage and autonomously activate the agent delivery. In this case AE is the most promising technique. Imitating the blood circulatory system, in the near future healing pipes will be three-dimensional, building a vascular supply network.

## Healing Efficiency Assessment

The test configuration used to quantify healing efficiency is not standardized. Imitating the service loads applied on load-bearing concrete elements in reality, a quasi-static bending test is selected. As a first approach, small-scale plain normal concrete beams are cast carrying a pre-notch at the bottom side of the middle section. Under bending, a unique crack initiates from the pre-notched area and propagates along the beam's height. The encapsulated healing material is positioned during casting at the zone above the pre-notch ensuring that the propagated crack will pass through the tubular capsule/pipeline. The aforementioned test setup is empirically



**Fig. 7** (a) Short tubular glass capsules filled with two-component healing agent and attached to concrete casting mold; (b) Short glass capsules attached to large-scale concrete beam mold; (c) Vascular system assembly: long tubular ceramic capsules connected to 3D printed reservoirs; (d) Network of Vascular Tubes attached to the formwork

chosen (Tsangouri 2015) and designed according to Rilem TC 50-FMC recommendation (Rilem TC 50-FMC 1985) in order to assess the healing efficiency in the idealized scenario at which a unique crack evolves and autonomously heals without the intervention of other developed cracks in the vicinity or any reinforcement restrictions. Bending test is applied in two cycles: in the first cycle, concrete pre-cracks, capsules break releasing the agent and activating the repair process and in the second test cycle (after agent curing) bending is repeated under the same circumstances in order to assess the healed crack response to damage. The repair performance is quantified after healing by measuring the strength, stiffness recovery, and the residual fracture toughness (Tsangouri 2015).

More realistic test configurations consider small or full-scale rebars reinforced concrete beams or slabs that include encapsulated healing system (De Belie et al. 2015; Van Tittelboom 2016). Under quasi-static four-point bending, multiple cracks form and interact. In this case, the healing system should provide efficient and

repeatable repair at different locations. Strength regain is not an indicator of efficient healing since rebars plastic deformation mask potential concrete strength restoration. Stiffness and fracture toughness measurement is also affected by the presence of rebars that continue to slip, debond, or deform in the second test cycle. Healing efficiency assessment remains challenging and cannot be based on mechanical feature restoration. AE appears a promising alternative tool and was used during the last decade in this direction. An overview of the main findings is given at the following paragraphs highlighting the most promising analysis tools.

## Test Setup

During this series, small- and large-scale concrete samples were tested in quasi-static mode under three- and four-point bending. Bending stopped when the crack width was measured equal to 300 µm (in multiple cracks case, the average of cracks width is considered) (Tsangouri 2015). A loading pause of 24 h followed at which the healing agent released after cracking was allowed to cure. The loading was repeated following the same procedure in order to assess the healing efficiency. The two-cycles loading/reloading test was repeated several times to assess the repeatability of healing.

## AE Setup

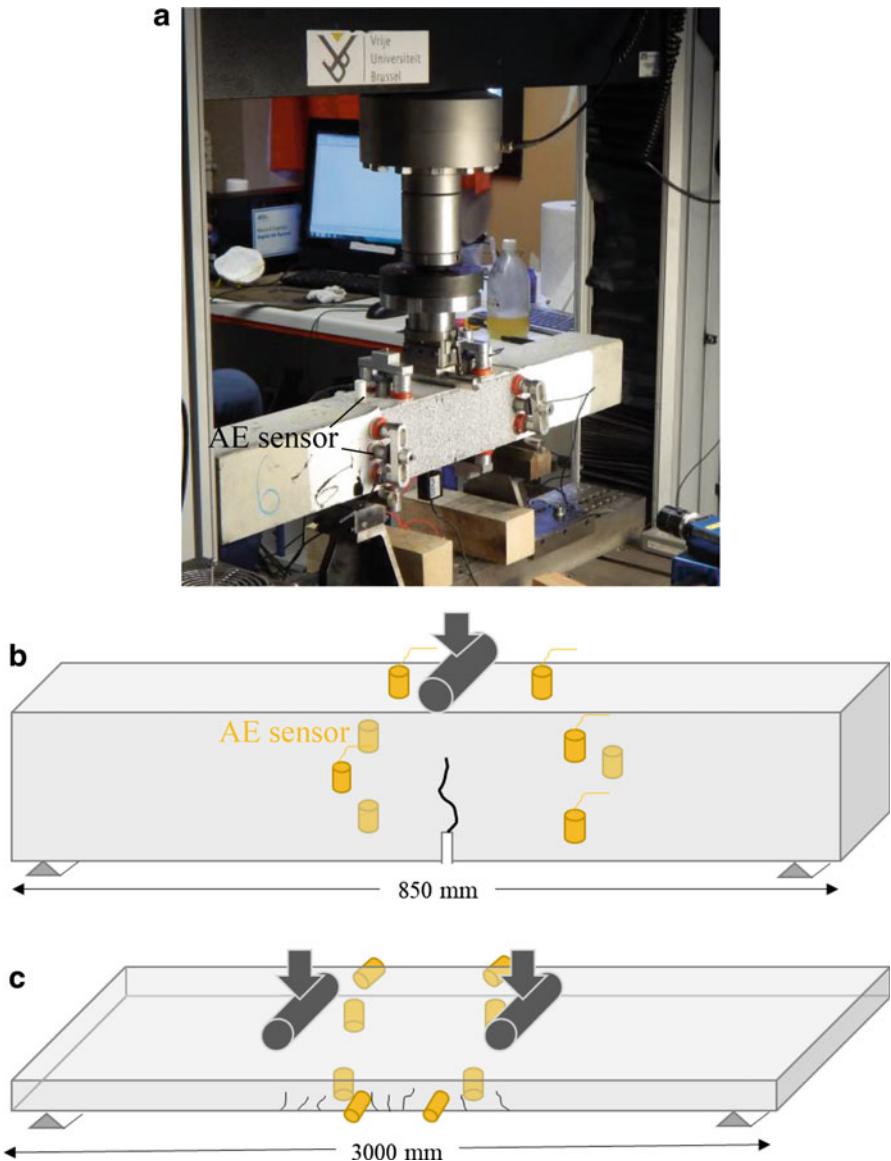
An array of eight resonant R15 (resonance frequency 150 kHz) sensors is mounted on the concrete surface using Vaseline coupling agent and supported by magnetic holders (Fig. 8a). The sensors are positioned surrounding the area where damage was expected to develop: around the pre-notch and the constant-moment zone in the case of three- and four-point bending test, respectively (further details in Tsangouri et al. 2013a; Van Tittelboom et al. 2016; Lelon 2017). In the drawings in Fig. 8b, c, two representative sensor setups are given (Table 2).

The threshold was set at 40 dB. The AE features and the waveform shape were stored using the AEwin software. The AE source was localized using 3D triangulation algorithm. The transition wave velocity and attenuation were measured in advance using pencil lead break test (Hsu-Nielsen source) at 4000 m/s and 15 dB/m

## Results

### Detecting Capsule Breakage Using AE Features Analysis

Accurate capsule breakage detection is important since it indicates the damage triggering and need for healing activation. AE events due to glass capsule breakage that provide high energy hits, are first differentiated from concrete cracking which releases lower energy-carrying hits (Tsangouri et al. 2013a). Based on that ascertainment, a series of tests assessed capsule breakage based on AE energy release.



**Fig. 8** (a) AE sensors attached on concrete beam tested under 3-point bending; (b) AE sensors position for small-scale beams; and (c) large-scale slab series, respectively

In Fig. 9, an overview of results obtained on tests of small-(a-b) and large-scale (c) concrete beams carrying short glass capsules is presented. It is shown that when a glass tube breaks, a high-energy elastic wave is released. As a result, each AE sensor placed at the vicinity of the event capture instantly a high energy hit. It should be noted that the magnitude of signals emitted due to capsule breakage is significantly

**Table 2** Self-healing samples design and test setup overview

Encapsulation approach	Dimensions (in mm)	Test setup	Crack pattern
Short tubular glass/ceramic/cement capsules (Fig. 7a)	Small-scale concrete beam 850 × 100 × 100 mm <sup>3</sup> Small-scale mortar beam 160 × 40 × 40 mm <sup>3</sup>	3-point bend Pre-notched	Unique crack initiated from the pre-notch
Short tubular glass capsules (Fig. 7b)	Large-scale beam 3000 × 400 × 300 mm <sup>3</sup>	4-point bending No pre-notch	Stochastic multiple cracks
2D tubular ceramic pipes + internal agent reservoir (Fig. 7c)	Small-scale beam 850 × 100 × 100 mm <sup>3</sup>	3-point bend Pre-notched	Unique crack initiated from the pre-notch
2D tubular ceramic pipes + external agent reservoir (Fig. 7d)	Large-scale slab 3000 × 1000 × 250 mm <sup>3</sup>	4-point bending No pre-notch	Stochastic multiple cracks

greater than the rest of signals: energy, duration, and counts are up to one order of magnitude higher in this case. Discussion in this section focuses only on energy feature since it is the most representative and indicative AE feature (Tsangouri et al. 2013a). The discrete events of capsule breakage are well differentiated from the rest of the AE activity. The range of energy depends on the capsule material and the local fracture process as well as the sensor-source distance and the sample geometry.

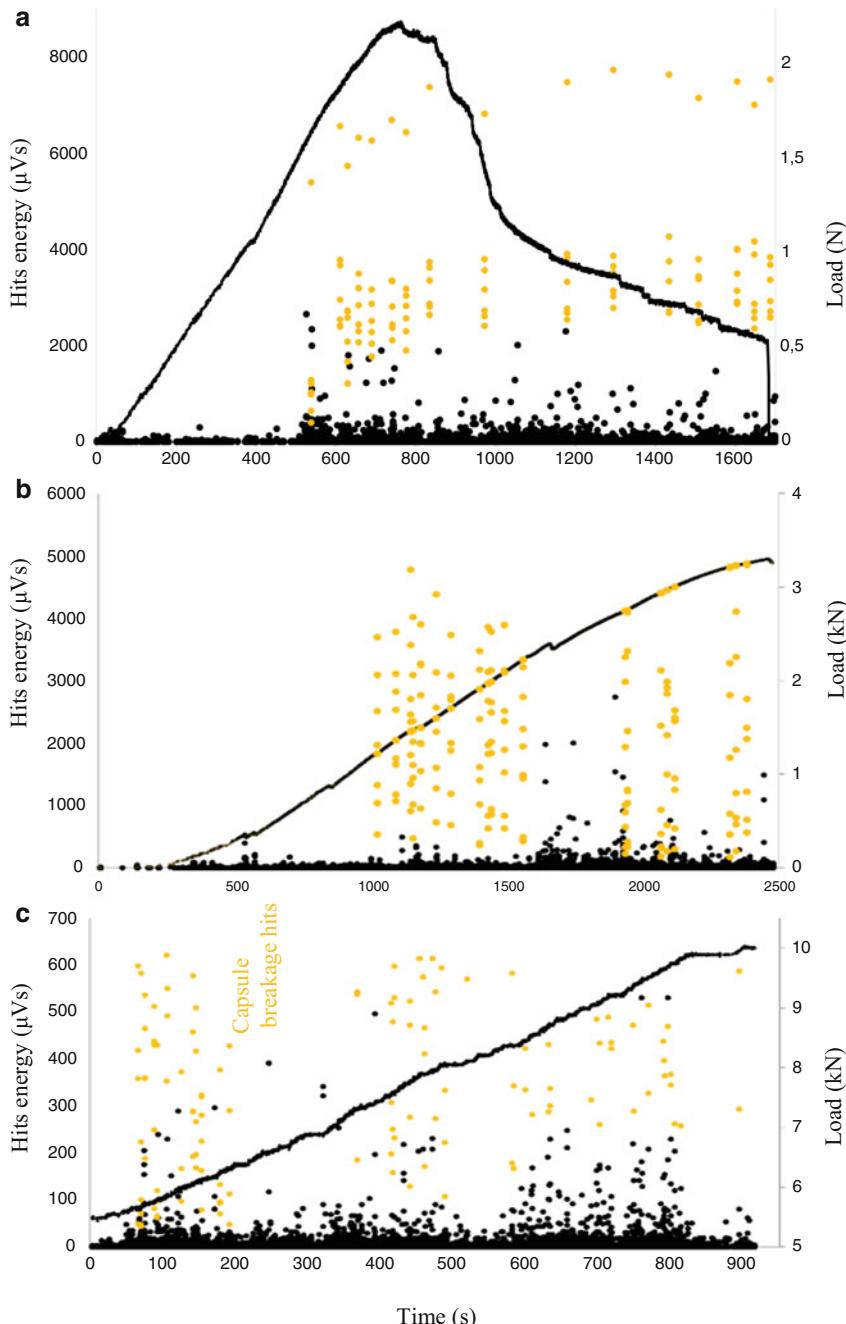
In detail, as presented in Fig. 9a, the instant of capsule breakage, the sensors placed closer to the capsule receive a signal with greater energy than the signals captured by sensors located further away. For this reason, a typical event of capsule breakage consists of hits with energy range values (measured in  $\mu$ Vs) from 12,000 to a few hundreds. The signal shape and features also change in the presence of cracks (single crack (Fig. 9a) (Tsangouri et al. 2013a) or multiple cracks (Fig. 9b, c) (Van Tittelboom et al. 2015, 2016)) that distort and scatter the wave released before reaching the sensors located on the surface.

The capsule-concrete bonding properties affect the way the capsule breaks (Gilabert et al. 2017). It is found that glass capsules exhibit weak bonding with concrete, therefore their brittle breakage is accompanied by debonding of the capsule from the surrounding concrete. On the contrary, studies at which cementitious or ceramic capsules are used have shown that AE energy levels of capsule breakage hits were lower but still can be differentiated from concrete cracking hits energy (Minnebo et al. 2017).

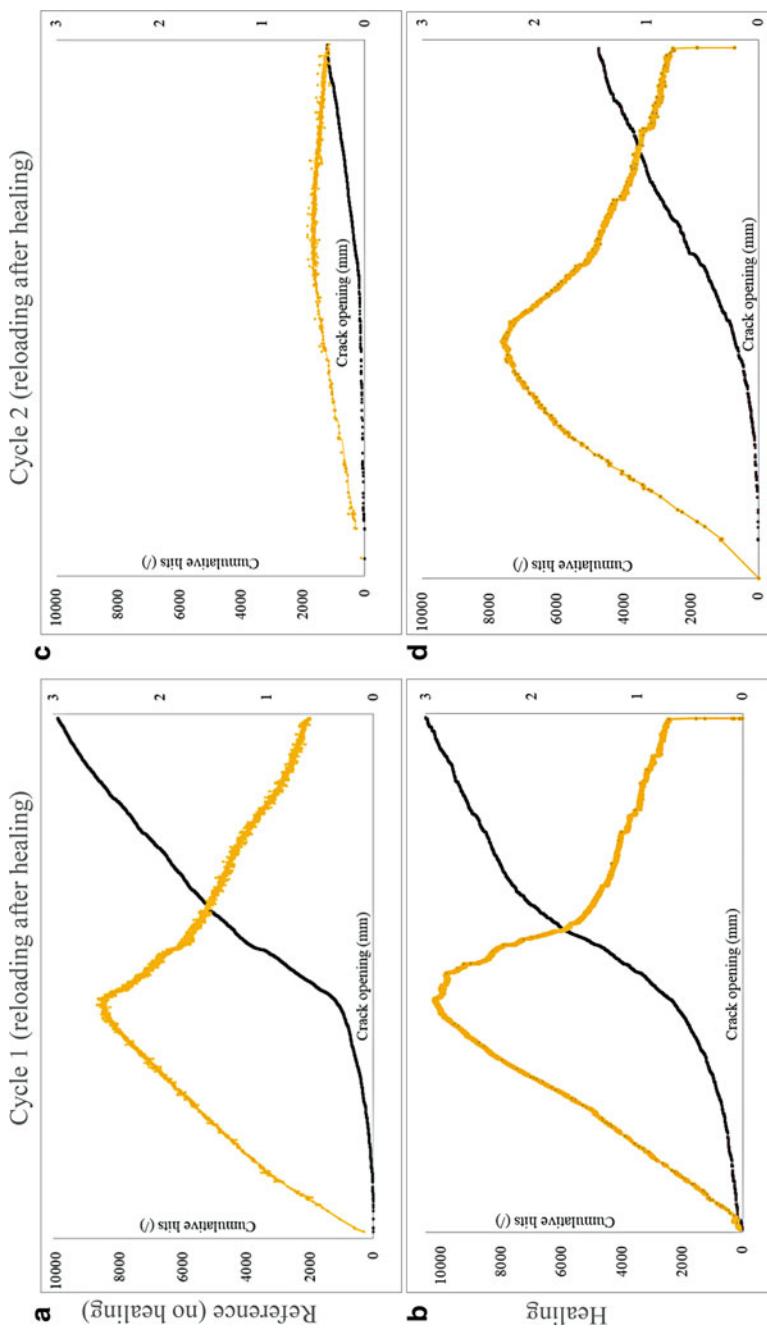
### Hits Distribution Assessing Damage Processes and Healing Agent Breakage

As previously discussed, cumulative hits distribution can be a good indication of fracture process in concrete. From that perspective in this section, concrete damage and repair after healing are analyzed.

**Concrete damage.** To begin with, in Fig. 10a, the hits evolution as a single crack forms at the pre-notched section of concrete beams tested under bending is



**Fig. 9** Glass capsule breakages determined using AE hits energy distribution: (a) small-scale concrete beam tested under 3-point bending where a unique crack is developed; (b) small-scale concrete beam tested under 4-point bending where multiple cracks form; (c) large-scale concrete beam tested under 4-point bending where multiple cracks form

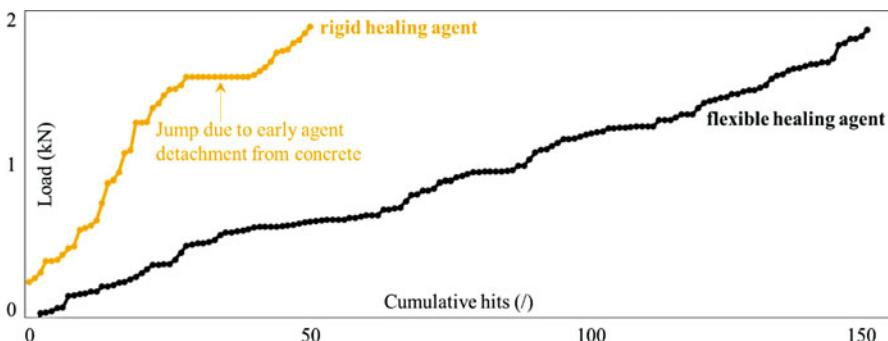


**Fig. 10** Cumulative hits distribution and load evolution in the case of (a, b) reference and (c, d) healing series at first (crack formation) and second (after healing, crack reopening) loading cycle, respectively

presented. The load evolution is given in the secondary, vertical axis. As shown, almost negligible AE hits occur at the initial linear loading stage. As soon as the linearity is lost, the macro-crack forms and the number of hits increases significantly. Beyond the peak load, hits rate increases further as the capsules rupture and the macro-crack opens. The three stages of hits evolution discussed in previous section are clearly shown in this graph. Similar to Fig. 10a, in Fig. 10b, the same test procedure is applied, but now the beam was reinforced with glass capsules that break in the presence of macrocracks during testing. Most of the capsule breakage events occur at the post peak stage and slightly change the cumulative hits slope.

**Healing of cracks.** After capsule breakage, healing agent is released into the crack void and concrete is autonomously repaired. At a second loading cycle, the hits distribution is given for the reference (no healing) and the healed concrete samples in Fig. 10c, d, respectively. It is shown that the number of AE hits at the reference case was significantly lower indicating that the crack reopens and widens further at this second loading cycle without any noticeable new damage activity. On the other hand, there was AE activity at the reloading cycle of the healed sample. The cumulative hits curve has the pattern of the previous test cycle. This is an indication that the crack has been repaired and resists further damage, similar to healthy samples (Tsangouri et al. 2014).

In a recent study AE hits analysis was considered in order to select the healing agent that can optimally bridge moving cracks in concrete (Feiteira et al. 2017). Rigid (strong, with elastic modulus of 22 MPa) and flexible (less strong, with elastic modulus of 10 MPa) polyurethane-based foams were encapsulated into glass capsules and embedded into small-scale mortar beams that crack under bending. The AE activity after cracks healing was measured indicating that the rigid healing agent developed several jumps (vertical discontinuities) at the accumulated hits evolution and several high energy (one order of magnitude above most AE events) hits were detected correlated to early detachment of the agent from the cement matrix. This is not the case for the flexible healing agent that plastically deforms reaching large elongation, therefore hits released are low in energy and cumulative hits gradually evolve (Fig. 11).



**Fig. 11** Cumulative hits distribution during bending test after crack healing considering rigid (yellow color) and flexible (black color) healing agent

## Source Localization Locating Agent Release Areas and New Cracks Formed after Healing

In Fig. 12a, the localized events detected during testing on a small-scale concrete beam under bending are projected in plane. The events emitted due to fracture process zone building up surrounding the macrocrack are colored in yellow. The events (selected by AE hits energy analysis, see previous section) emitted as capsules instantly break are colored in black. It is shown that AE 3D source localization algorithms can precisely detect capsule breakages. The latter is a significant property of AE analysis since knowing the exact position of capsule rupture, the zone of released agent can be located, therefore the zone where repair is expected.

In Fig. 12b, the projection map of capsule breakage events at the crack plane is presented. In gray and black color, the actual and AE calculated locations of the capsules are, respectively, given. The variation between the real and the calculated positions is less than 8 mm constituting good engineering accuracy (Tsangouri et al. 2013a).

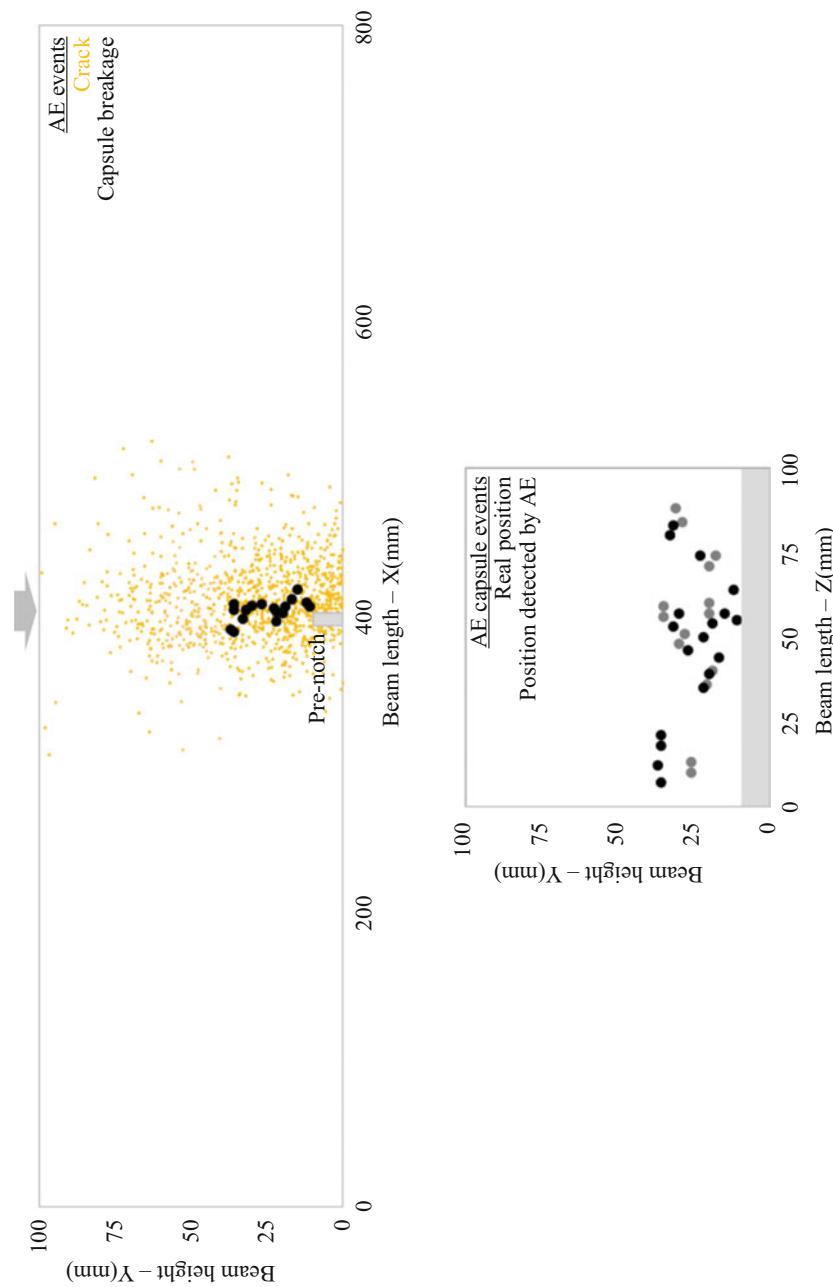
AE localization analysis provides evidence of successful healing in concrete by detecting cracks closure and new crack formation after healing. In Fig. 13a, b, two representative cases are given: concrete beams are loaded under bending and autonomously healed by agent that travels into the vascular network reaching the crack void. In Fig. 13a, two macro-cracks form at the zone close to the 3D printed reservoirs (highlighted in black dots). At reloading stage, after healing, the crack at the right is fully restored and cannot reopen. Another crack forms in the vicinity (yellow-colored AE events). Respectively, in Fig. 13b, the healed crack does not reopen after healing (black-colored AE events at the left side). It is shown that simple AE events distribution projected on a plane (beam's side view) indicates efficient concrete repair (Tsangouri 2015).

---

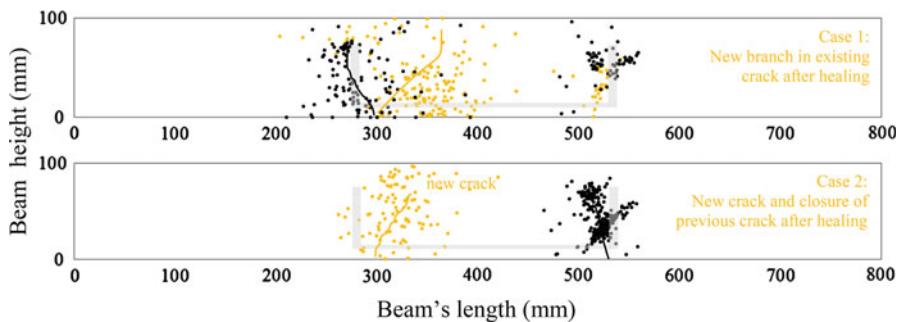
## Discussion and Future Perspectives

### Summary

This chapter presents the Acoustic Emission NDE technique principles, the instrumentation and the methodology applied to assess materials health and damage conditions. The study focuses on the application of AE on concrete structures and especially discusses the case study of a recently developed concrete technology, namely autonomously healed and repaired concrete. A series of most promising AE analysis tools are extensively presented. The performance of the AE technique is evaluated and the main contribution of the method toward the damage characterization is discussed. The case study concerns quite a complex concrete geometry involving small- and large-scale concrete samples that carry encapsulated healing agent or embedded vascular network pipelines connected to 3D-printed agent reservoirs. AE cumulative hits distribution appears as a promising monitoring tool



**Fig. 12** (a) Projection map of AE events emitted due to concrete cracking (yellow) and capsule rupture (black); (b) Real (gray) and AE localized (black) position of capsules at the crack plane



**Fig. 13** AE events projected on a plane detected in a small-scale concrete beam including 2D tubular ceramic pipeline-reservoir healing system at first (black) and second (yellow) bending test cycle; (a) case in which a new branch is formed after healing at the existing crack; (b) case in which a crack is repaired after healing and does not open, instead a new crack forms at a different position

that associates AE activity with the structural integrity. AE source localization based on simple triangulation algorithms accurately locates the damage onset and source position. AE wave features are employed to detect the healing activation (capsules rupture), distinguish the damage mechanisms, and evaluate the mechanical restoration after healing.

It is shown that an array of sensors attached on concrete surface provides essential information that fully characterizes materials' response to damage. AE is classified as an NDE technique since even early and limited AE activity can be indicative of material mechanical performance. In other words, tests on limited stresses (nondestructive) can be used to predict the remaining service life of materials and constructions.

### Problem of Signal Complexity

The main advantage of the AE technique, namely that signals are generated not artificially but from the material itself can be challenging regarding results interpretation. Considering that the AE sensor is attached to the material surface, an acoustic wave emitted at the interior must travel through the material in order to reach the surface. As the wave travels through solids, it distorts, attenuates, and scatters. Consequently, AE waveform captured by AE sensors often carries the propagating signal from the source along with reflections. Respectively, waves travelling through solids suffer from attenuation affects: as the propagation distance increases, the wave amplitude weakens. For the aforementioned reasons, the received signal is often a continuous signal with features that significantly differ from the burst signal emitted from the AE source. One should be cautious and identify the numerous wave propagation phenomena that make practical implementation of AE difficult.

## Synergy with Other NDE Techniques

Even if the use of AE as the monitoring method can provide a full-field view of damage, the combination with other NDEs is recommended in order to further and in-depth assess the severity of damage. Optical and other acoustic methods are commonly chosen to accompany the AE technique. The list of techniques is quite long: ultrasound pulse velocity, optical microscopy, computer tomography scanning, digital image correlation, and shearography. Briefly, representative studies are discussed: crack closure after healing is detected by AE cumulative hits distribution and visualized by digital image correlation crack opening fields (Feiteira et al. 2017; Tsangouri et al. 2013b). Ultrasound pulse velocity measurements were performed applying sensors on concrete surface (Malm and Grosse 2016) or embedded into concrete (Tsangouri et al. 2015) providing accurate damage qualification.

## Future Perspectives

Today is an era of revolutionary technological advances that lead to new durable, strong, and sustainable construction materials. Robust composites are designed with thinner and lighter sections. The evaluation of structural integrity of these innovative materials requires new AE instrumentation. Wired AE hardware that transfer signals from each sensor to the recording device will be soon replaced by a network of wireless sensors and contactless recording devices (Grosse and Krüger 2006). The AE of the future will be able to provide robust mobile long-term monitoring of structures.

Especially in the case of highly attenuated concrete structures, there is a significant need toward the embedment of AE sensors into concrete (during casting). This way the damage propagation path is limited and the diffusion effects due to surface reflections are eliminated. Preliminary works show that piezoelectric sensors embedment into concrete is feasible (Tsangouri et al. 2015). However, further research is required in order to build up embedded sensor networks that monitor the structural health of large-scale concrete elements.

From another perspective, computational studies that use AE measurements as input to accurately assess damage zones in concrete are gaining ground. For instance, AE events detected during inspection can be used for wave velocity mapping base on Travel Time Tomography (Behnia et al. 2014). The computational tomographic reconstruction can precisely identify the progressive damage accumulation even in the case of large-scale complex concrete structures.

Taking into account the recent advances in AE studies, it is evident that monitoring the evolution of simple AE features (energy, RA, frequencies, etc.) can indicate damage onset and propagation. Rilem Technical Committees (e.g., TC 212-ACD) are active in this direction aiming to build up test recommendations that combine AE with other NDE techniques in order to detect and evaluate damage in concrete.

In parallel, the effect of propagation distance on waveform analysis should be clarified. It is proven that more accurate AE studies can be obtained only if the emissions are classified according to the sensor-source distance (signals travelling

greater distances carry dominant reflections and attenuated/distorted wave components) (Tsangouri et al. 2016).

## Advantages and Limitations of AE

In Table 3, the main advantages and limitations of the AE NDE technique are summarized.

**Table 3** Advantages and limitations of AE technique

Advantages	Limitations
Application at different stages of structure life: from proof of testing at design stage to structural health assessment during service life	Extraneous noise: preset threshold and signal filter required
Global monitoring of structures using only a sensor array	Free surface required to attach sensors
Real-time damage evaluation	Good surface coupling required
High sensitivity	Particular results not reproducible due to signal source nature
Accurate source localization without point-to-point scanning	Signal shape affected by material and geometry attenuation

### Glossary

Acoustic Emission	Physical phenomenon whereby transient elastic waves are emitted due to internal material movements
Amplitude (dB or V)	The maximum (peak) value that a waveform attains
Attenuation (dB/m)	The effect of signal magnitude drops as the distance of sensor from the source increases
Average Frequency (kHz)	The ratio of threshold crossings (counts) over the signal duration
Count (/)	Number of times that the signal exceeds the amplitude threshold
Duration ( $\mu$ s)	Time interval between the signal onset (first threshold crossing) and the last threshold crossing
Energy ( $\mu$ Vs)	Area under the waveform envelope corresponding to the energy contained in the signal
Event (/)	A local material change giving rise to acoustic emission
Felicity effect	The presence of detectable acoustic emission at fixed sensitivity level, below previously exceeded stress levels
Hit (/)	Detection and identification of signal by an AE sensor
Kaiser effect	The absence of detectable acoustic emission at a fixed sensitivity level, until previously applied stress levels are exceeded

Source localization	Computing method that detects source location and time based on localization triangulation algorithm that considers the signal arrival times of the sensors array and a predetermined wave propagation velocity
Rise time (μs)	Time interval between the signal onset and the peak amplitude
RA value (μs/V)	The ratio of signal rise time over its amplitude
Sensor	Detection apparatus made by piezoelectric PZT plate that senses transient elastic waves and transforms it into electrical signal
Signal	The electrical signal coming from the transducing element and passing through the subsequent signal conditioning element
Stress wave	Wave emitted as material responds to any type of stressor
Threshold (dB)	Amplitude level set such that only signals with amplitude greater than this level are identified and stored

## References

- Aggelis DG (2011) Classification of cracking mode in concrete by acoustic emission parameters. *Mech Res Commun* 38(3):153–157. <https://doi.org/10.1016/j.mechrescom.2011.03.007>
- Aggelis DG, Dassios KG, Kordatos EZ, Matikas TE (2013) Damage accumulation in cyclically-loaded glass-ceramic matrix composites monitored by acoustic emission. *Sci World J* 869467:10. <https://doi.org/10.1155/2013/869467>
- Behnia A, Chai HK, Yorikawa M, Momoki S, Terazawa M, Shiotani T (2014) Integrated non-destructive assessment of concrete structures under flexure by acoustic emission and travel time tomography. *Constr Build Mater* 67:202–215. <https://doi.org/10.1016/j.conbuildmat.2014.05.011>
- Carpinteri A, Lacidogna G, Paggi M (2007) Acoustic emission monitoring and numerical modeling of FRP delamination in RC beams with non-rectangular cross-section. *Mater Struct* 40(6): 553–566. <https://doi.org/10.1617/s11527-006-9162-4>
- De Belie N, Van Tittelboom K, Tsangouri E, Karaikos G, Snoeck D, Wang J, Araujo A, Van Hemelrijck D (2015) Autonomous regeneration of concrete structures by incorporation of self-healing mechanisms. In: International conference on the regeneration and conservation of concrete structures (RCCS), Nagasaki
- Dry C, McMillan W (1996) Three-part methylmethacrylate adhesive system as an internal delivery system for smart responsive concrete. *Smart Mater Struct* 5(3):297. <https://doi.org/10.1088/0964-1726/5/3/007>
- Dunegan HL, Harris DO, Tatro CA (1968) Fracture analysis by use of acoustic emission. *Eng Fract Mech* 1(1):105–122. [https://doi.org/10.1016/0013-7944\(68\)90018-0](https://doi.org/10.1016/0013-7944(68)90018-0)
- Evans AG, Clifton JR, Anderson E (1976) The fracture mechanics of mortars. *Cem Concr Res* 6(4):535–547. [https://doi.org/10.1016/0008-8846\(76\)90082-X](https://doi.org/10.1016/0008-8846(76)90082-X)
- Feiteira J, Tsangouri E, Gruyaert E, Lors G, Louis G, De Belie N (2017) Monitoring crack movement in polymer-based self-healing concrete through digital image correlation, acoustic emission analysis and SEM in-situ loading. *Mater Des* 115:238–246. <https://doi.org/10.1016/j.matdes.2016.11.050>

- Ghosh SK (ed) (2009) Self-healing materials: fundamentals, design strategies and applications. Wiley-VCH, Weinheim. <https://doi.org/10.1002/9783527625376>
- Gilabert F, Van Tittelboom K, Tsangouri E, Van Hemelrijck D, De Belie N, Van Paepegem W (2017) Determination of strength and debonding energy of a glass-concrete interface for encapsulation-based self-healing concrete. *Cem Concr Compos* 79:76–93. <https://doi.org/10.1016/j.cemconcomp.2017.01.011>
- Granger S, Loukili A, Pijaudier-Cabot G, Chanvillard G (2007) Experimental characterization of the self-healing of cracks in an ultra high performance cementitious material: mechanical tests and acoustic emission analysis. *Cem Concr Res* 37(4):519–527. <https://doi.org/10.1016/j.cemconres.2006.12.005>
- Grosse C, Krüger M (2006) Wireless acoustic emission sensor networks for structural health monitoring in civil engineering. In: European conference on non-destructive testing (ECNDT), Berlin
- Grosse C, Malm F (2016) Combination of inspection and monitoring techniques for the detection of fractures in concrete with self-healing properties. In: 9th international conference on fracture mechanics of concrete and concrete structures, FraMCoS-9
- Grosse C, Reinhardt H, Dahm T (1997) Localization and classification of fracture types in concrete with quantitative acoustic emission measurement techniques. *NDT E Int* 30(4):223–230. [https://doi.org/10.1016/S0963-8695\(96\)00060-6](https://doi.org/10.1016/S0963-8695(96)00060-6)
- Kaiser J (1953) Erkenntnisse und Folgerungen aus der Messung von Gerad'uschen bei Zugbeanspruchung metallischen Werkstoffen (in German). *Steel Res Int* 24(1–2):43–45
- Karaikos G, Tsangouri E, Aggelis DG, Daraemaeker A, Van Hemelrijck D (2016) Performance monitoring of large-scale autonomously healed concrete beams under four-point bending through multiple non-destructive testing methods. *Smart Mater Struct* 25(5):055003. <https://doi.org/10.1088/0964-1726/25/5/055003>
- Kobayashi AS, Hawkins NM, Chan YA, Lin IJ (1980) A feasibility study of detecting reinforcing-bar debonding by acoustic-emission technique. *Exp Mech* 20(9):301–308. <https://doi.org/10.1007/BF02328623>
- Landis EN (1999) Micro–macro fracture relationships and acoustic emissions in concrete. *Constr Build Mater* 13(1):65–72. [https://doi.org/10.1016/S0950-0618\(99\)00009-4](https://doi.org/10.1016/S0950-0618(99)00009-4)
- Lelon J (2017) Experimental testing to assess the upscaling of vascular self-healing on concrete. Master dissertation, Vrije Universiteit Brussel, Brussels
- Maji A, Shah SP (1988) Process zone and acoustic-emission measurements in concrete. *Exp Mech* 28(1):27–33. <https://doi.org/10.1007/BF02328992>
- Malm F, Grosse C (2016) Efficiency of self-healing agents for cementitious materials characterized by NDT. World Conference on NDT (WCNDT 2016), Munich
- Minnebo P, Thierens G, De Valck G, Van Tittelboom K, De Belie N, Van Hemelrijck D, Tsangouri E (2017) A novel design of autonomously healed concrete: towards a vascular healing network. *Materials* 10(1):49. <https://doi.org/10.3390/ma10010049>
- Mirmiran A, Philip S (2000) Comparison of acoustic emission activity in steel-reinforced and FRP-reinforced concrete beams. *Constr Build Mater* 14(6):299–310. [https://doi.org/10.1016/S0950-0618\(00\)00036-2](https://doi.org/10.1016/S0950-0618(00)00036-2)
- Mobasher B, Stang H, Shah SP (1990) Microcracking in fiber reinforced concrete. *Cem Concr Res* 20(5):665–676. [https://doi.org/10.1016/0008-8846\(90\)90001-E](https://doi.org/10.1016/0008-8846(90)90001-E)
- Nielsen J, Griffin DF (1977) Acoustic emission of plain concrete. *J Test Eval* 5(6):476–483. <https://doi.org/10.1520/JTE10561J>
- Ohtsu M (1987) Acoustic emission characteristics in concrete and diagnostic applications. *JAE* 6(2):99–108
- Otsuka K, Date H (2000) Fracture process zone in concrete tension specimen. *Eng Fract Mech* 65(2):111–131
- Pollock A (1969) Stress-wave emission in ndt. *Non-Destr Test* 2(3):178–182. [https://doi.org/10.1016/0029-1021\(69\)90109-1](https://doi.org/10.1016/0029-1021(69)90109-1)
- Rilem TC 212-ACD (2010) Acoustic emission and related NDE techniques for crack detection and damage evaluation in concrete. *Mater Struct* 43(9):1183–1186. <https://doi.org/10.1617/s11527-010-9639-z>

- Rilem TC 50-FCM (1985) Determination of fracture energy of mortar and concrete by means of three-point bend tests on notched beams. *Mater Struct* 18(106):285–290. <https://doi.org/10.1007/BF02498757>
- Rossi P, Robert JL, Gervais JP, Bruhat D (1989) Identification of the physical mechanisms underlying acoustic emissions during the cracking of concrete. *Mater Struct* 22(3):194–198
- Rossi P, Robert JL, Gervais JP, Bruhat D (1990) The use of acoustic emission in fracture mechanics applied to concrete. *Eng Fract Mech* 35(4–5):751–763. [https://doi.org/10.1016/0013-7944\(90\)90158-D](https://doi.org/10.1016/0013-7944(90)90158-D)
- Seo Y, Kim YR (2008) Using acoustic emission to monitor fatigue damage and healing in asphalt concrete. *KSCE J Civ Eng* 12(4):237–243. <https://doi.org/10.1007/s12205-008-0237-3>
- Shah SP, Choi S (1999) Nondestructive techniques for studying fracture processes in concrete. *Int J Fract* 98(3):351–359. <https://doi.org/10.1023/A:1018620008780>
- Shield CK (1997) Comparison of acoustic emission activity in reinforced and prestressed concrete beams under bending. *Constr Build Mater* 11(3):189–194. [https://doi.org/10.1016/S0950-0618\(97\)00036-6](https://doi.org/10.1016/S0950-0618(97)00036-6)
- Shiotani T, Momoki S, Chai H, Aggelis DG (2009) Elastic wave validation of large concrete structures repaired by means of cement grouting. *Constr Build Mater* 23(7):2647–2652. <https://doi.org/10.1016/j.conbuildmat.2009.01.005>
- Shull PJ (ed) (2016) Nondestructive evaluation: theory, techniques, and applications. Marcel Dekker, Inc. New York, Basel
- Thierens G, De Valck G (2015) A novel design of autonomously healed concrete-towards the agent vascular network. Master dissertation, Vrije Universiteit Brussel, Brussels
- Tsangouri E (2015) Experimental assessment of fracture and autonomous healing of concrete and polymer systems. PhD dissertation, Vrije Universiteit Brussel (VUB), Brussels
- Tsangouri E, Aggelis DG, Van Tittelboom K, De Belie N, Van Hemelrijck D (2013a) Detecting the activation of a self-healing mechanism in concrete by acoustic emission and digital image correlation. *Sci World J* 424560:10. <https://doi.org/10.1155/2013/424560>
- Tsangouri E, Van Tittelboom K, Van Hemelrijck D, De Belie N (2013b) Visualization of the healing process on reinforced concrete beams by application of digital image correlation (DIC). *Mater Charact VI* 77:283. <https://doi.org/10.2495/MC130251>
- Tsangouri E, Karaikos G, Aggelis D, Deraemaeker A, Van Hemelrijck D (2014) Healing performance on concrete under mode I fracture by ultrasonic testing using embedded transducers, acoustic emission and digital image correlation method. In: 9th International conference on structural dynamics (EURODYN 2014), Porto
- Tsangouri E, Karaikos G, Aggelis DG, Deraemaeker A, Van Hemelrijck D (2015) Crack sealing and damage recovery monitoring of a concrete healing system using embedded piezoelectric transducers. *Struct Health Monit* 14(5):462–474. <https://doi.org/10.1177/1475921715596219>
- Tsangouri E, Karaikos G, Deraemaeker A, Van Hemelrijck D, Aggelis D (2016) Assessment of acoustic emission localization accuracy on damaged and healed concrete. *Constr Build Mater* 129:163–171. <https://doi.org/10.1016/j.conbuildmat.2016.10.104>
- Van Tittelboom K (2012) Self-healing concrete through incorporation of encapsulated bacteria-or polymer-based healing agents. PhD dissertation, University of Ghent, Ghent
- Van Tittelboom K, De Belie N, Lehmann F, Grosse C (2012) Acoustic emission analysis for the quantification of autonomous crack healing in concrete. *Constr Build Mater* 28(1):333–341. <https://doi.org/10.1016/j.conbuildmat.2011.08.079>
- Van Tittelboom K, Tsangouri E, Van Hemelrijck D, De Belie N (2015) The efficiency of self-healing concrete using alternative manufacturing procedures and more realistic crack patterns. *Cem Concr Compos* 57:142–152. <https://doi.org/10.1016/j.cemconcomp.2014.12.002>
- Van Tittelboom K, Wang J, Araújo M, Snoeck D, Gruyaert E, Debbaut B, Derluyn H, Cnudde V, Tsangouri E, Van Hemelrijck D, De Belie N (2016) Comparison of different approaches for self-healing concrete in a large-scale lab test. *Constr Build Mater* 107:125–137. <https://doi.org/10.1016/j.conbuildmat.2015.12.186>



# Guided Wave Testing

5

Paul Fromme

## Contents

Introduction .....	142
Guided Wave Propagation .....	143
Guided Wave Testing of Plate Structures .....	147
Defect Detection in Isotropic (Metal) Plates .....	147
Multilayered Plates, Stiffeners, and Lap-Joints .....	151
Anisotropic Structures, Composites .....	153
Curved Structures .....	156
Pipes .....	156
Experimental Methods .....	158
Simulations .....	162
Summary .....	163
References .....	164

## Abstract

Guided waves can propagate long distances in thin-walled structures, such as pipelines or plates. This allows for the efficient monitoring and testing of large structures and for the detection of hidden or inaccessible defects. Guided wave propagation is dispersive and multi-modal, requiring a thorough understanding of the wave propagation and scattering phenomena from simulations. Guided wave dispersion diagrams, mode shapes, and typical signals are illustrated for the example of isotropic plates. Both low and high frequency guided waves have been used for the testing of plate structures, with different wave modes and applications including tomography and arrays for the

---

P. Fromme (✉)

Department of Mechanical Engineering, UCL, London, UK

e-mail: [p.fromme@ucl.ac.uk](mailto:p.fromme@ucl.ac.uk)

detection and localization of defects. For multilayered and anisotropic structures, guided wave propagation becomes more complex, and often the fundamental guided wave modes are employed for defect detection. For pipelines different commercially available testing systems have been developed and long propagation distances up to 100 m have been achieved. Careful selection of guided wave mode and excitation frequency allows the minimization of attenuation due to viscoelastic coatings and in buried pipelines. Synthetic focusing using non-axisymmetric modes improves defect imaging and localization. Experimental methods differ from standard ultrasonic testing, as good control of the excited guided wave mode shape and signal are required to achieve improved sensitivity for small defects. In addition to contact piezoelectric transducers, electromagnetic and laser techniques allow for non-contact measurements. Finite Element Analysis is one of the numerical simulation techniques used to obtain a better understanding of guided wave testing and to improve defect characterization.

---

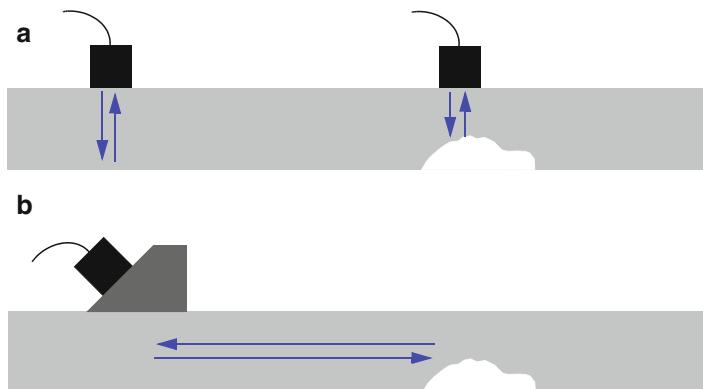
## Introduction

Increasingly guided waves are employed for nondestructive testing (NDT) and structural health monitoring (SHM) applications, as they offer capabilities for the testing of large structures and detection of hidden or inaccessible defects. For structures that are thin in at least one dimension, such as plates and pipes, through multiple reflections in the thickness direction, standing wave modes can develop that can propagate over long distances with limited amplitude loss (Cawley et al. 2003). Commercially successful equipment has been developed by a number of manufacturers for the testing of pipes (Rose 2002a; Mudge 2001; Alleyne et al. 2017), where defects can be detected at distances up to 100 m.

Compared to conventional bulk ultrasonic testing, the wave propagation direction is along the structure rather than through the thickness (Fig. 1), reducing the requirement for scanning and offering the potential to inspect structures with limited access, such as buried or coated pipelines (Rose 2014). Increasingly, permanently installed sensors are employed, allowing a monitoring rather than inspection approach to reduce variations between successive measurements and improve the sensitivity for defect detection.

However, there is typically a balance between the ability to cover a large area from a single sensor location and the sensitivity for the detection of small defects. Applications for large structures require lower frequencies to achieve long wave propagation distances, with the large wavelength corresponding to a reduced sensitivity for small defects. Often guided waves are used in a screening mode to identify potentially damaged areas and provide a preliminary assessment of the damage severity, rather than a detailed defect characterization and sizing.

While providing many benefits, it must be understood that guided wave propagation is more complex than for bulk ultrasonic waves and depends both on the



**Fig. 1** Schematic of (a) bulk ultrasonic and (b) guided wave testing

geometry and excitation frequency. Multiple wave modes can propagate in the structure and the propagation velocity depends on the excitation frequency. This dispersion and the resulting pulse distortion require a good understanding of the wave propagation and control of the wave excitation to ensure reliable defect detection. This has been recognized by the international norms for guided wave testing (e.g., BS 9690–1:2011, ASTM E2775–16) and necessitates improved training to provide reliable testing results.

This chapter aims to provide an overview of guided waves for nondestructive testing. Section “[Guided Wave Propagation](#)” explains some of the important phenomena in more detail (for the example of plates) to provide a better understanding of the application, benefits, and limitations of guided wave testing. Application examples are discussed in section “[Guided Wave Testing of Plate Structures](#)” for isotropic and anisotropic (composite) plates, as well as bonded, multilayered, and curved structures. Section “[Pipes](#)” gives an overview of applications for pipes, with a focus on how the choices for wave modes influence both the defect detection sensitivity and experimental and processing requirements. Sections “[Experimental Methods](#)” and “[Simulations](#)” provide a brief discussion of experimental and simulation methods for guided wave testing.

---

## Guided Wave Propagation

Guided waves can propagate along plates, with multiple modes and varying directionality. Conceptually, this can be understood either as the constructive interference of multiple bulk ultrasonic waves reflected and mode converted through the thickness of the structure, or as guided wave modes through the thickness. Both conceptual derivations result in identical equations, which for the case of a plane wave front in an isotropic, homogeneous plate can be described by the classical Rayleigh-Lamb theory.

$$\frac{\tan(qh)}{\tan(ph)} = -\frac{4k^2pq}{(q^2 - k^2)^2} \text{ Symmetric Lamb wave modes} \quad (1)$$

$$\frac{\tan(qh)}{\tan(ph)} = -\frac{(q^2 - k^2)^2}{4k^2pq} \text{ Anti-symmetric Lamb wave modes} \quad (2)$$

where  $p^2 = \left(\frac{\omega}{c_L}\right)^2 - k^2$  and  $q^2 = \left(\frac{\omega}{c_T}\right)^2 - k^2$  with plate thickness  $2h$ , angular frequency  $\omega$ , wavenumber  $k$ , longitudinal wave velocity  $c_L$ , and shear wave velocity  $c_T$ .

These equations and the wave propagation have been widely described and discussed in the seminal work by Achenbach (1973), Auld (1973), Graff (1975), Rose (2014), and Viktorov (1967). For the simple case of a uniform plate, an analytical solution is possible. For more complex structures, numerical methods such as SAFE (semi-analytical finite element) analysis (Predoi et al. 2007) or semi-analytical approaches such as the global matrix approach employed in the Disperse software (Pavlakovic et al. 1997) have been demonstrated to be numerically stable and efficient. Figure 2 shows the dispersion curves for an aluminum plate calculated using the Disperse software.

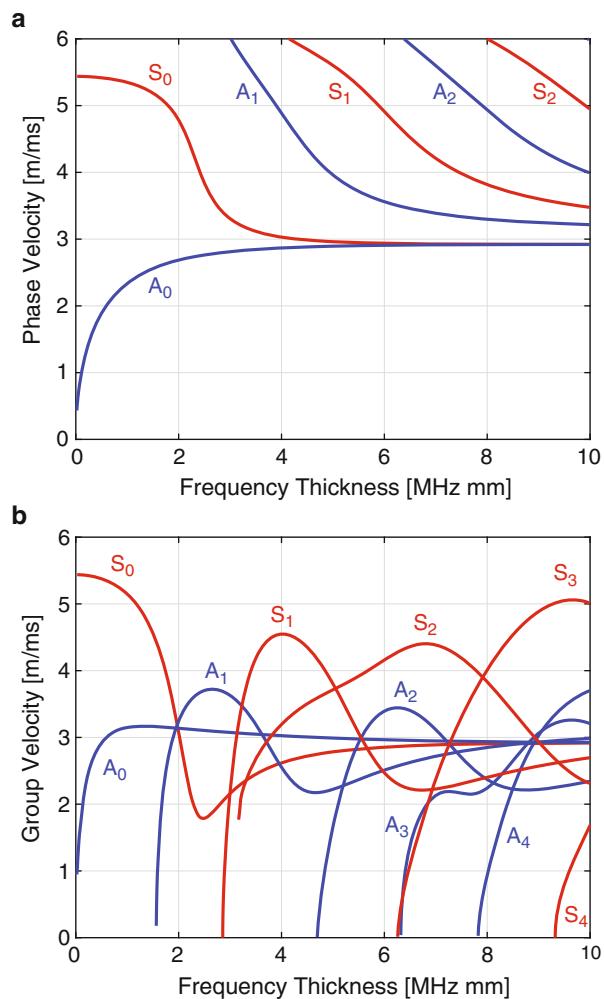
Figure 2a shows the phase velocity for the different wave modes for variation of the frequency-thickness product, as the wave speed depends on this product. The corresponding group velocities (the speed of the wave pulse) are shown in Fig. 2b. For higher frequency-thickness products, multiple wave modes can propagate, above the cut-off frequencies of the higher wave modes.

At the cut-off frequencies the respective group velocity of that wave mode tends towards zero, while the phase velocity tends towards infinity. Essentially, at the cut-off frequency this can be thought of as a standing wave mode or vibration across the plate thickness. Below the cut-off frequencies of the higher wave modes, only the two fundamental wave modes can propagate. The first anti-symmetric wave mode  $A_0$  can be considered similar to a bending or flexural wave at low frequency-thickness, with large out-of-plane displacement and smaller, anti-symmetric in-plane motion, as shown in Fig. 3a. At low frequencies, the first symmetric Lamb mode  $S_0$  resembles a longitudinal wave with large, symmetric in-plane displacement (Fig. 3b) and very limited dispersion, i.e., the phase and group velocity are reasonably constant and similar to a bulk, longitudinal ultrasonic wave.

For high frequency-thickness products, the fundamental  $S_0$  and  $A_0$  Lamb wave modes tend towards the Rayleigh surface wave velocity with similar mode shapes through the thickness (Fig. 3c), while the higher Lamb wave modes (Fig. 3d) modes tend towards the shear velocity.

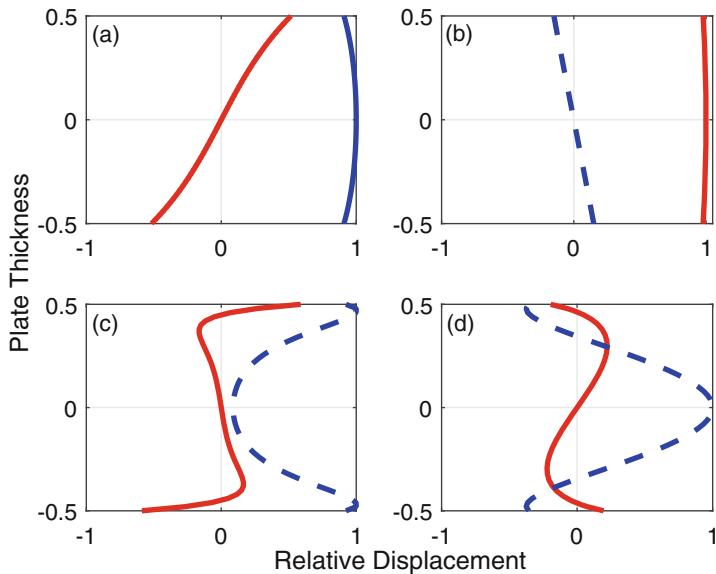
The complexity of the guided wave propagation requires different considerations as compared to conventional bulk ultrasonic wave testing. Dispersion, the fact that different frequency components within a wave packet propagate at different velocities, leads to changes in the pulse shape and makes some of the standard UT evaluation (e.g., based on pulse shape) difficult. Furthermore, the

**Fig. 2** Dispersion diagram for aluminum plate: (a) phase velocity; (b) group velocity

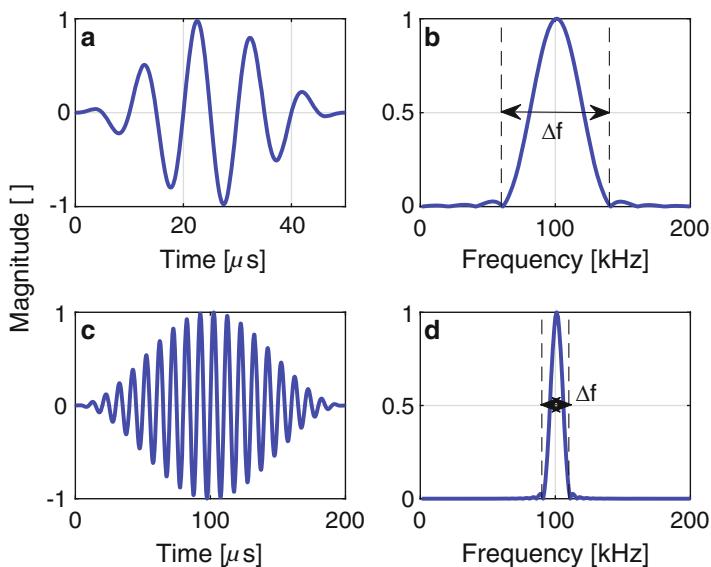


variation in group velocity with frequency leads to pulse distortion, the spreading in time of the energy contained within a wave pulse. This reduction in amplitude with propagation distance was one of the most relevant problems encountered during early stages of guided wave testing (Alleyne and Cawley 1991). Instead of the conventional short duration, wide frequency range pulse employed in standard ultrasonic testing in combination with resonant transducers, different excitation signals are often employed for guided wave testing.

In order to achieve a narrow frequency bandwidth to limit dispersion and pulse distortion, tone bursts such as five cycles of a sinusoid in a Hanning window have been found beneficial (Fig. 4a, b).



**Fig. 3** Guided wave mode shapes; **(a)**  $A_0$  mode at 0.5 MHz mm; **(b)**  $S_0$  mode at 0.5 MHz mm; **(c)**  $A_0$  mode at 10 MHz mm; **(d)**  $A_1$  mode at 10 MHz mm; solid: in-plane displacement; dashed: out-of-plane displacement



**Fig. 4** Five cycle sinusoid (100 kHz center frequency) in Hanning window: **(a)** time signal, **(b)** frequency content, 20 cycle sinusoid in Hanning window, **(c)** time signal, **(d)** frequency content

$$\Delta f = \frac{4}{N}f \text{ with center frequency } f \quad (3)$$

The number of cycles  $N$  controls the trade-off between pulse length and frequency bandwidth  $\Delta f$  (Eq. 3), so a 20 cycle tone burst (Fig. 4c, d) has longer time duration but narrower frequency bandwidth than a five cycle pulse (Fig. 4a, b). Another approach has been to use a wideband excitation such as a chirp signal, and to post-process the measured time traces by convolution with narrowband signals to isolate specific frequencies within the bandwidth of the excitation (Michaels et al. 2013). In order to achieve this control of the excitation signal, typically different equipment compared to standard ultrasonic pulser/receivers is required, such as arbitrary function generators to generate the desired time signals, and wide-band amplifiers to achieve acceptable signal-to-noise (SNR) ratios. This will be discussed further in section “[Experimental Methods](#)” dealing with experimental methods.

Another important consideration is that signals containing multiple wave modes propagating at different speeds can be quite difficult to analyze (Alleyne and Cawley 1991), thus often selective excitation of Lamb wave modes is desired. This is reasonably straight-forward to achieve for the fundamental wave modes below the cut-off frequencies of the higher wave modes. The  $S_0$  wave mode has been widely employed as it is non-dispersive at low frequencies, its mode shape and displacement are similar to longitudinal bulk ultrasonic waves (Fig. 3b), and it has the highest group velocity, thus is easily identified as the first arrival pulse. The wave mode has some limited dispersion around 0.5 MHz mm, which has been used for tomography to detect and localize plate thinning (Leonard et al. 2002). Disadvantages of the  $S_0$  mode at low frequency thickness products are the rather large wavelength and limited sensitivity for some defects in composite structures (Guo and Cawley 1993), but compared to the  $A_0$  mode, it has the advantage of limited attenuation for water loading. In contrast the  $A_0$  wave mode is dispersive at low frequencies, but limited pulse distortion was found around 0.5 MHz mm for steel and aluminum plates. Together with the shorter wavelength, this allows for good defect detection sensitivity (Fromme et al. 2006). At higher frequency-thickness products, the selective mode excitation requires better experimental configuration and good understanding of the guided wave propagation and attenuation characteristics.

## **Guided Wave Testing of Plate Structures**

### **Defect Detection in Isotropic (Metal) Plates**

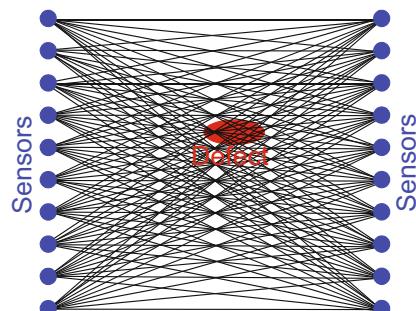
Guided waves can quickly interrogate large plate and shell structures and have sensitivity for the detection of surface and subsurface features (Rose 2002b). Guided waves propagate in two dimensions along thin structures, making it necessary to achieve a distinction between the different Lamb wave modes and to take the two-dimensional spreading of the guided ultrasonic wave into account. The angular

resolution of any monitoring device has to be sufficient to distinguish between different features. The amplitude of the guided wave decreases with distance, as the energy spreads over a larger area. Therefore, a sufficiently large dynamic range is required for the detection of small defects in the presence of structural features. In order to simplify the signal processing, most nondestructive testing applications work in the low frequency-thickness regime below the cut-off frequencies of the higher wave modes, where only three guided wave modes ( $S_0$ ,  $A_0$ ,  $SH_0$ ) can exist.

For metallic structures, the most commonly investigated defect types include corrosion and fatigue cracking. Applications include the thinning of plates due to corrosion, e.g., at difficult to reach areas such as pipeline supports. Tomography approaches rely on transducers located around the area of interest and changes in the wave pulse arrival time due to thickness reduction (Fig. 5). This requires a dispersive guided wave mode, where the velocity changes with thickness in the frequency range of interest, but not too significantly to completely alter the wave propagation characteristics. In particular, the  $S_0$  mode in a frequency-thickness range below 2 MHz mm has been used for tomography applications (Malyarenko and Hinders 2000), as at low frequency-thickness product it has the highest group velocity, i.e., the fastest arrival time. For single defects of simple geometry, a good match of the estimated and actual thickness reduction was found (Leonard et al. 2002). For more complex or deep defect geometries or sharp edges, additional effects due to scattering have to be considered and can make accurate sizing challenging. Different reconstruction algorithms and guided wave modes have been investigated for their suitability and compared for specific thickness reductions (Zhao et al. 2011). Higher wave modes with shorter wavelength due to the higher excitation frequency offer better sensitivity for shallow defects, but less accurate depth sizing of deep defects was found. In general, good results for the detection and localization of corrosion thinning using guided wave tomography have been reported, but accurate sizing can be more complicated depending on the defect type and severity.

Significant work has been reported on the first symmetrical Lamb wave mode  $S_0$  and the lowest shear mode  $SH_0$ , as their displacement, stress distribution, and propagation velocity at low frequencies are similar to the bulk ultrasonic wave modes and non-dispersive (Zhao and Rose 2004b). The wave propagation characteristics are similar to the torsional and longitudinal modes often employed for pipe

**Fig. 5** Schematic of tomography on plate, sensors, wave paths, and defect marked



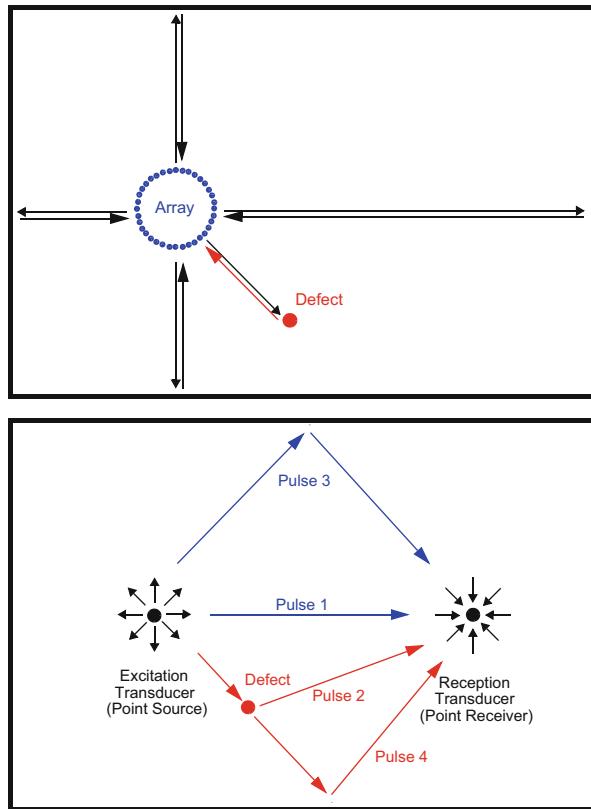
applications, and for thin wall thickness compared to the pipe diameter locally similar defect sensitivity was observed. The scattering at part and through thickness notches and cracks oriented along and perpendicular to the wave propagation direction was studied experimentally and from FE simulations. Significant variations depending on the defect depth and length relative to the wavelength were found (Demma et al. 2003).

The interactions with damage as well as the appropriate damage monitoring strategy were investigated (Lee and Staszewski 2007). For fatigue crack detection, the scattering at cracks emanating from circular cavities (holes) was investigated. The scattering at holes in a plate was studied analytically for flexural waves ( $A_0$  mode) (Pao and Chao 1964; Fromme and Sayir 2002b), and for the  $S_0$  mode (McKeon and Hinders 1999; Diligent et al. 2002), to establish a baseline. (Fromme and Sayir 2002a) investigated the changes to the scattered field during fatigue crack growth and found good agreement of experimental results with numerical predictions. Guided waves were employed to successfully detect defects at difficult to reach locations around a hole (Doherty and Chiu 2012) and fatigue cracks emanating at fasteners within a lap joint (Cho and Lissenden 2012). The scattering behavior of guided waves at complete and partial through-thickness notches (Lowe and Diligent 2002; Fromme and Rouge 2011) and cracks (Chang and Mal 1999) has been predicted and validated experimentally.

As for aerospace structures the areas of high stress concentration are typically known, higher frequency guided waves offer a potential trade-off between smaller monitoring range and improved sensitivity for small defects. Higher frequency guided wave modes have been increasingly used for improved defect detection over shorter distances, as the ratio of the wavelength to the defect size determines the detection sensitivity. Compared to the lower frequency, fundamental Lamb modes, experimentally the selective excitation of specific modes is more complicated but allows for easier interpretation of results. Finite element simulations combined with modal decomposition were used to study the interaction of Lamb waves with defects for frequency-thickness products up to 5 MHz mm (Terrien et al. 2007). High frequency Lamb waves at about 15 MHz mm and the application for crack detection were demonstrated (Greve et al. 2008). The detection of fatigue cracks and other defects using the high frequency fundamental Lamb wave modes (Chan et al. 2015; Masserey and Fromme 2017), the  $A_1$  mode (Khalili and Cawley 2016), and higher order mode clusters (HOMC) (Ratnam et al. 2012) were demonstrated.

For flat structures, guided wave arrays (Fig. 6a) have been employed (Li and Rose 2001), (Salas and Cesnik 2009). Wilcox et al. (2005) developed a compact, moveable guided wave phased array system using electromagnetic acoustics transducers (EMATs) to excite and receive the  $S_0$  mode. With post-processing similar to conventional ultrasonic phased arrays or in the wavenumber-frequency domain (Wilcox 2003), defects in a metallic plate could be localized. Similar concepts were developed by Fromme et al. (2006) for the  $A_0$  mode, but employing a circular array of bonded piezoelectric sensors. Different types of defects such as partial and through holes and plate thinning similar to corrosion were successfully detected. The

**Fig. 6** (a) Schematic of localized array; (b) schematic pulses distributed array



influence of the number and pattern of the sensors on the sidebands and thus selectivity for defect detection were studied (Wilcox et al. 2005; Yu and Giurgiutiu 2012). The sensitivity for defect detection has been improved by employing different signal processing methods (Velichko and Wilcox 2008).

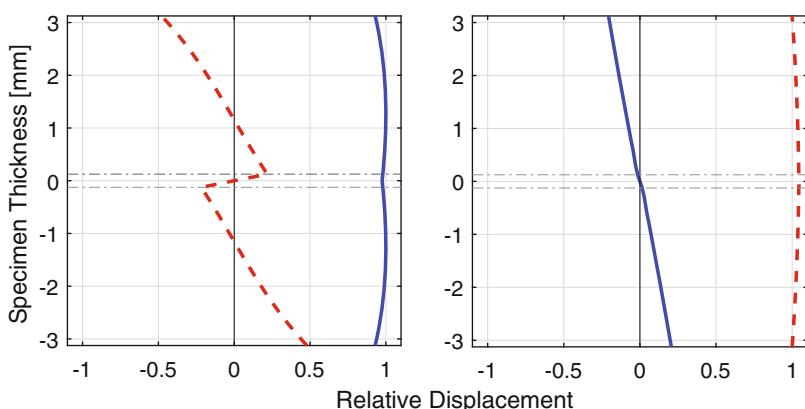
As localized phased array type transducers have limited sensitivity for defects (cracks) aligned radially to the sensor location, distributed guided wave sensor arrays (Fig. 6b) were investigated (Wang et al. 2004; Michaels 2008). Such approaches using permanently bonded sensors are at the cross-over between NDE and Structural Health Monitoring (SHM), often also called Structural Integrity Monitoring (SIM). The main advantage of permanently attached transducers is the reduction of variability if a baseline subtraction approach is employed, which is often necessary to distinguish between reflections at structural features and developing defects, especially for more complex structures (Attarian et al. 2014). Damage localization is achieved using the differenced signals (baseline subtraction) to separate scattered signals from structural features and damage. Different signal processing algorithms have been implemented and tested, ranging from elliptical imaging (Wang et al. 2004) to more advanced methods, such as minimum variance imaging (Hall et al. 2014), maximum-likelihood estimation (Flynn et al. 2011), and sparse

reconstruction (Levine and Michaels 2013). Guided wave imaging algorithms have the potential for damage characterization, as the defect geometry influences the scattering pattern, which can be taken into account to determine likely defect shape and orientation, e.g., for minimum variance imaging (Hall et al. 2014).

However, environmental effects such as temperature changes, surface conditions, and loading can affect the arrival time of different signal components and thus the accuracy of the baseline subtraction (Konstantinidis et al. 2006; Lu and Michaels 2009; Chen et al. 2012), limiting the effective dynamic range and thus sensitivity. Improvements can be achieved with more advanced signal processing algorithms or corrections for benevolent variations of the baseline, e.g., optimum baseline subtraction and baseline stretching (Croxford et al. 2010).

### Multilayered Plates, Stiffeners, and Lap-Joints

For complex structures consisting of multiple metallic layers, often connected using sealant layers and fasteners for aerospace applications, or bonded stiffeners and lap joints, the wave propagation and scattering characteristics are typically more complex than for single metallic layers. Taking as an example a structure consisting of two aluminum layers with a sealant layer as a simplified model for aircraft structures, the guided wave mode shapes through the thickness become more complex. Similarities to single layered structures exist, but the fundamental guided waves (not strictly Lamb waves) at low frequencies now consist of the individual bending of the metallic layers (Kostson and Fromme 2009), either in-phase to give a dominant out-of-plane displacement similar to the  $A_0$  Lamb mode or out-of-phase with a predominant in-plane displacement (Fig. 7). Higher-order guided wave mode shapes tend to



**Fig. 7** Mode shapes for multilayer structure (3 mm aluminum, 0.25 mm epoxy, 3 mm aluminum), 100 kHz frequency; **(a)** anti-symmetric mode; **(b)** symmetric mode; solid: out-of-plane; dashed: in-plane displacement

be even more complex and depend significantly on the frequency, geometry, and material properties.

Lower frequency guided wave modes in multilayered structures can be considered in analogy to single metal plates and good propagation distances have been achieved (Dalton et al. 2001), as the attenuation due to the sealant layer is typical at acceptable levels. As the guided wave modes have energy distributed through the complete specimen thickness, defect detection is possible even in areas with lack of sealant, but the wave propagation characteristics are influenced by the local contact conditions. The flexural wave mode has been employed to detect and monitor fatigue crack growth adjacent to a fastener hole (Kostson and Fromme 2009), and innovative approaches such as instrumented fasteners (Rakow and Chang 2012) have been proposed. Suitable modes of high frequency guided waves for the inspection of different layers of multilayered structures can be found employing a mode-tuning technique (Quarry 2004). For multilayered aircraft structures high frequency guided wave modes have been shown to be sensitive for the detection of manufactured notches (Lindgren et al. 2007), and fatigue cracks (Chan et al. 2015), but propagation distances are limited due to attenuation and a proper understanding of the wave propagation characteristics is required.

Bonded stiffeners are widely employed in aerospace structures, contributing in a complex assembly with the aircraft skin and the airframe, to provide strength and stiffness. Adhesive bonding achieves a continuous bond and results in good stiffness and force transfer with relatively low stresses. The quality of the adhesive bonds between aircraft skin and stiffener are important for the structural performance (Di Scalea et al. 2007). Problems include poor adhesion and inadequate strength of the adhesive material caused by incorrect curing (Thompson and Thompson 1991). The velocity of bulk ultrasonic waves in the adhesive material, which depends to a large degree on material stiffness, may be used to monitor curing of the adhesive (Rokhlin et al. 1981; Freemantle and Challis 1998; Dixon et al. 2004).

Guided waves can be used to monitor continuous bond lines, if their propagation is sensitive to the properties of the bond. This idea has been employed for adhesive joints of parallel plates using Lamb and interface waves (Mal et al. 1989; Nagy and Adler 1989). Different possible guided wave modes are reviewed in (Lowe and Cawley 1994). The adhesive thickness and properties influence the guided wave propagation characteristics. Viscoelastic adhesives contribute to the attenuation of the higher order guided wave modes in multilayered plate structures (Seifried et al. 2002). Nagy and Adler (1989) used guided waves to inspect adhesive joints between plates. Guided waves propagating in a plate structure can be used in a different configuration to interrogate attachments, e.g., spar on skin (Di Scalea et al. 2007) or repair patch on skin (Le Crom and Castaings 2010; Puthillath and Rose 2010). Wave modes guided along a structural feature such as a weld or bonded stiffener have been considered. Trapped modes can exist, e.g., due to the geometry leading to a lower phase velocity than in the surrounding plate, concentrating energy at the feature and thus enabling long distance propagation. This has been demonstrated for the case of a butt weld between two plates (local thickness increase) for a partially trapped compression-like mode (Sargent 2006) and a perfectly trapped shear-like mode with

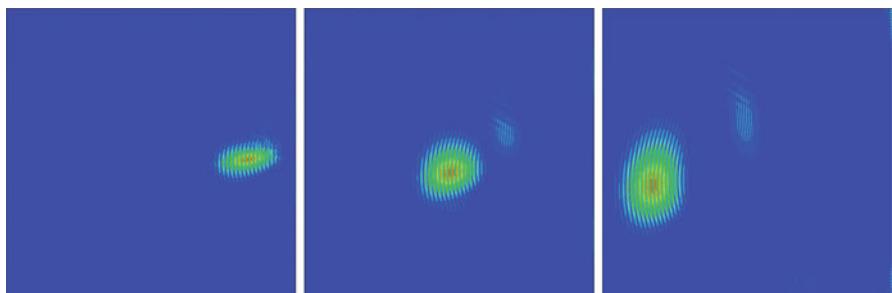
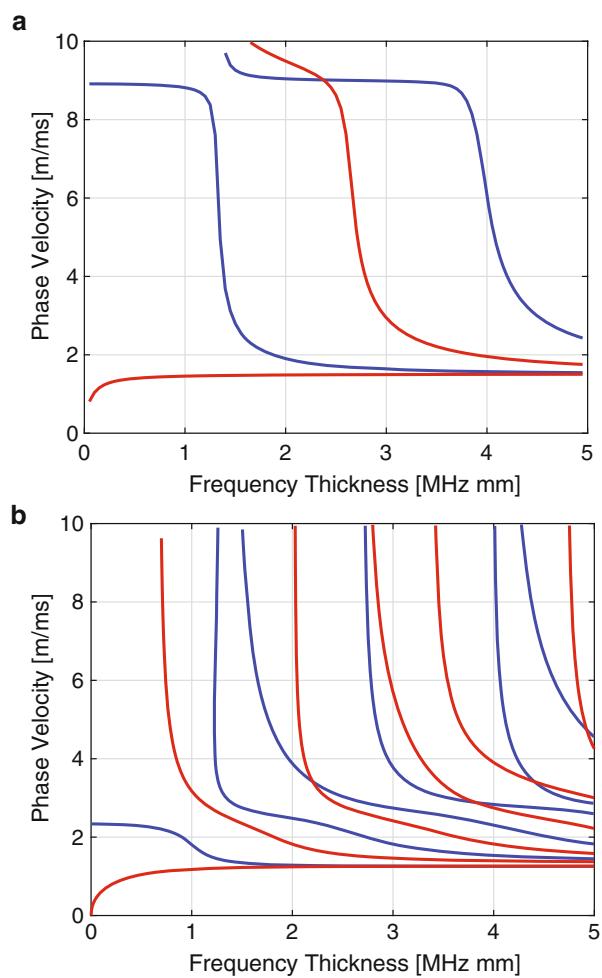
very little dispersion (Fan and Lowe 2009; Castaings and Lowe 2008). These wave modes can be employed to inspect the weld material and the material in the heat-affected zone (HAZ). Shear modes propagating along a stiffener have been used to monitor the curing of the bond line to a plate structure (Fan et al. 2013).

## Anisotropic Structures, Composites

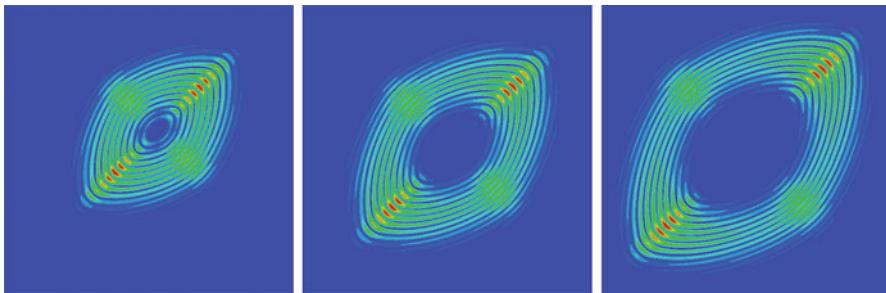
Anisotropic and inhomogeneous materials are widely employed, e.g., carbon or glass fiber-reinforced pre-preg composites for aircraft structures and wind turbine blades, as they offer improved strength to weight ratios. The degree of anisotropy depends strongly on the material, e.g., unidirectional composites have stronger anisotropy than cross-ply or quasi-isotropic lay-ups. Typically material properties are less well known than for metallic materials and might vary within the structure due to manufacturing imperfections. Guided wave propagation characteristics can be predicted using a global matrix approach such as implemented in the Disperse software (Pavlakovic et al. 1997) or a transfer matrix approach (Hosten and Castaings 1993). The anisotropy has a strong effect on the  $S_0$  guided wave mode, as its velocity depends on the in-plane stiffness. For example, for a unidirectional plate, the  $S_0$  phase and group velocities in the fiber direction can be a factor 3 higher than perpendicular to the fiber orientation. The effect on the anti-symmetric  $A_0$  mode is lower, as the flexural wave mode depends as well on the properties of the matrix material (Fig. 8).

In the non-principal directions, the  $S_0$  and  $SH_0$  wave modes are coupled and the  $SH_0$  mode can exhibit cusps, complicating the wave propagation. Furthermore, wave skewing in the non-principal directions occurs for all wave modes, depending on the degree of anisotropy (Fig. 9). The skew angle describes the different directions of the phase and energy (group) velocities and can be calculated from the anisotropic material properties. Energy focusing along the fiber directions occurs, which can further complicate measurements. Chapuis et al. (2010) investigated the energy radiation of Lamb waves in a thin fiber-reinforced composite plate. Numerical and experimental results showed a significant direction-dependent focusing effect of the Lamb modes. Higher amplitudes and higher propagation speed along the fiber direction of a unidirectional composite plate is shown in Fig. 10. Group velocity curves in thin anisotropic, carbon fiber-reinforced epoxy laminates were measured using a point-source point-receiver configuration and compared to theoretical curves (Veidt and Sachse 1994). The influence of the anisotropy on Zero Group Velocity (ZGV) Lamb modes was investigated in monocrystalline silicon wafers using a line laser source (Prada et al. 2009). For an incident ultrasonic beam on an anisotropic multilayered structure, Potel et al. (2005) demonstrated that the Lamb wave beam generated in the plate can deviate with respect to the sagittal plane of excitation towards the stiffer direction of the anisotropic structure. Leleux et al. (2013) developed a multi-element matrix ultrasonic probe to inspect large composite plate components in pulse-echo mode from one single position and detect delamination and impact damage.

**Fig. 8** Dispersion diagram for unidirectional composite plate; (a) wave propagation in fiber direction; (b) perpendicular to fiber direction



**Fig. 9** Finite element simulation of S<sub>0</sub> guided wave mode propagation in monocrystalline silicon (0.38 mm thickness, 5 MHz frequency); showing wave skew relative to horizontal direction of wave excitation at three time points



**Fig. 10** Finite Element simulation of  $A_0$  guided wave mode propagation in  $45^\circ$  unidirectional composite (3.6 mm thickness, 100 kHz frequency) at three time points

The propagation of guided waves is complicated due to the anisotropic and inhomogeneous properties of the composites (Castaings and Hosten 2003). Together with typically high attenuation values, this makes monitoring and inspection using higher guided wave modes difficult and only limited work has been reported (Su et al. 2006). In general, it has been found advantageous to operate with a single wave mode at low frequency in order to avoid complications in the signal analysis and high attenuation. The fundamental symmetric mode  $S_0$  at low frequency has limited dispersion and the fastest propagation velocity, but the velocity depends strongly on the propagation direction relative to the composite layup fiber direction and the  $S_0$  mode is typically coupled with the  $SH_0$  mode (Datta and Shah 2009). It was found that the  $S_0$  mode is not sensitive to delaminations between plies being under zero shear stress condition (Guo and Cawley 1993).

For carbon and glass fiber composite materials, the critical damage mechanisms differ from metals. Manufacturing defects such as porosity, in- and out-of-plane waviness, and fiber misalignment can be inherent even in new composite structure and might necessitate acceptance inspection before service. Important in-service defects include high and low velocity impact damage, which can lead to varying damage mechanisms such as matrix and fiber breakage, matrix cracks, and delamination. The scattering of guided waves at a composite delamination is complex (Ng and Veidt 2011). The fundamental anti-symmetric mode  $A_0$  has a shorter wavelength than the  $S_0$  mode (Grondel et al. 2002) and thus in principle better sensitivity for defect detection. Furthermore, the directionality of the wave propagation characteristics is significantly less dependent on the anisotropic material properties, leading to similar velocities in all directions for quasi-isotropic and cross-ply (0/90) layups (Datta and Shah 2009). The  $A_0$  guided wave mode has been employed to detect different types of damage, such as cracking, fatigue, and delaminations in composite structures (Castaings et al. 2012). The  $A_0$  mode tends to be more sensitive to delaminations than the  $S_0$  mode and can detect delaminations at any depth (Guy et al. 2003). Mode conversion from the  $A_0$  to  $S_0$  mode was observed when the guided wave interferes with the delamination boundaries (Ramadas et al. 2010), confirmed from experimental work (Kazys et al. 2006). Delaminations can in principle be located by estimating the propagation speed and time of flight from the

reflected signal. Separate reflections from the delamination edges appear when the delamination length increases (relative to the wavelength). Composites subjected to impact damage were investigated (Kundu et al. 2008; Diamanti et al. 2004). The scattered wave amplitude pattern around a delamination showed a large forward scattered wave relative to the reflected pulse (Ng and Veidt 2011). Mesnil et al. (2014) investigated the energy trapping in a delamination area. Guided wave approaches have been proposed for the rapid inspection of silicon wafers to detect small cracks that increase wafer breakage rates. Chakrapani et al. (2012) used air-coupled transducers in pitch-catch configuration to generate the fundamental antisymmetric Lamb wave mode  $A_0$  in thin mono- and polycrystalline silicon wafers and detect cracks.

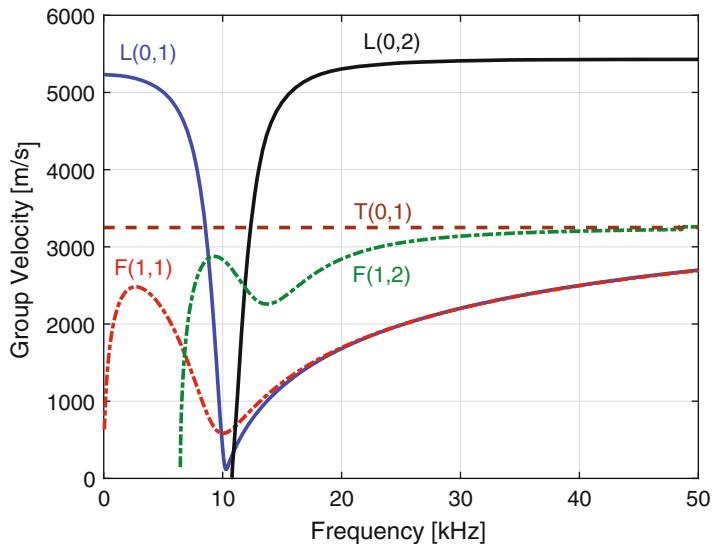
## Curved Structures

While the theoretical derivations above were developed for flat plates, in industrial applications often curved structures exist. The limitations of the applicability of concepts developed for flat plates has been investigated (Gridin et al. 2003), and it was found that the incurred error is typically low as long as the radius of curvature is significantly larger than the plate thickness. Therefore, most curved structures such as aircraft skins or tank walls can in good approximation be tested similar to plates. Within limitations this is also possible when localized wave propagation in large diameter, thin-walled pipelines is considered, e.g., around the circumference or along the length of the pipe (Liu and Qu 1998, Zhao and Rose 2004a). Circumferential guided waves have been employed to detect localized corrosion damage (Howard and Cegla 2017) or coating disbonds (Van Velsor et al. 2009).

---

## Pipes

Guided waves have been successfully employed for the nondestructive testing and monitoring of pipelines, with several commercially available testing equipment systems (Rose 2002a; Mudge 2001; Alleyne et al. 2017; Vinogradov et al. 2018). Several international norms have been established for guided wave testing of pipelines (e.g., BS 9690–2:2011, ASTM E2775–16, ASTM E2929–13). Pipes can be considered as one-dimensional structures along which the energy of axisymmetric guided ultrasonic wave modes propagates with little loss of energy, and propagation distances of up to 100 m have been achieved (Cawley et al. 2003). One of the main initial obstacles was the controlled excitation and reception of selected axisymmetric guided wave modes, as even at low frequency, multiple modes can be present. A typical dispersion diagram for a pipe is shown in Fig. 11. Modes are typically grouped as longitudinal (L), flexural (F), or torsional (T), with the number of nodes along the circumference and thickness labeled, e.g., the lowest longitudinal mode L (0,1) with uniform displacement around the circumference.



**Fig. 11** Dispersion diagram of group velocity for 6 inch, schedule 40 steel pipe; only lowest torsional mode T(0,1), flexural modes F(1,1) and F(1,2), and longitudinal modes L(0,1) and L(0,2) shown

Much of the early work employed the longitudinal mode L(0,2), as it has very limited dispersion and the fastest group velocity and thus earliest arrival time. Similarly, the torsional mode T(0,1) has been widely used for nondestructive testing, as it exhibits no dispersion and can be considered equivalent to the shear-horizontal (SH) mode in plates. For buried and coated pipes or liquid loading, attenuation increases depending on the viscosity and mode shape, limiting the achievable inspection distances and detection sensitivity. Viscoelastic coating is often used to insulate pipes and can lead to high attenuation, causing problems for guided wave testing. Care must be taken to select specific guided wave modes and excitation frequencies to obtain low attenuation and thus good propagation distances (Barshinger and Rose 2004). For pipes buried in sand, the attenuation increases depending on the excitation frequency, water saturation, and soil compaction (Leinov et al. 2015). Both the L(0,2) and T(0,1) mode have uniform stress and displacement around the circumference (no nodes) and the wave propagation can be considered as one-dimensional along the pipe, making analysis of the results and correlation to defect location along the pipe length straightforward. Typically, this is expressed as a percentage of the cross-sectional area (CSA) but allows only limited localization around the pipe circumference and classification of defect shape and type. Even for uncoated, above surface pipes, periodic welds and flanges between different sections of pipe act as reflectors and mostly uniform reflection around the circumference is received, which can be used to calibrate defect sensitivity. Depending on the employed guided wave mode and the geometry of the defect, e.g., circumferential or axial orientation of cracks, corrosion pitting, the reflection

and mode conversion depending on the extent, shape, and depth can be predicted (Cho et al. 1997; Bai et al. 2001; Ratassepp et al. 2010; Lovstad and Cawley 2012).

Synthetic focusing, based on post-processing of individual transducer signals, allows the improvement of defect imaging and localization. Different approaches for guided wave focusing were developed (Hayashi et al. 2005; Davies and Cawley 2009). The propagation of guided waves around pipe elbows poses a problem to achieve monitoring of the complete pipe volume at the bend and beyond. For detection of corrosion and erosion, especially the outside region of the bend can be difficult to monitor, as waves tend to travel the shortest distance around the inside of the bend. Monitoring systems (Brath et al. 2017) and focusing of non-axisymmetric modes (Rose et al. 2005) have been proposed, together with modeling to correct for pulse distortion (Sanderson et al. 2013). Long-term monitoring using permanently fixed electromagnetic acoustic transducers (EMAT) can help to improve the sensitivity for small defects (Herdovics and Cegla 2018).

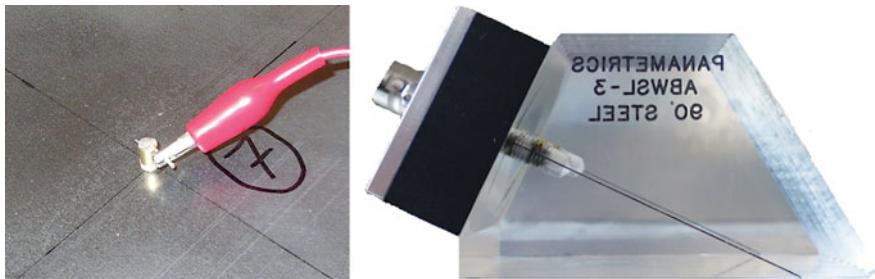
Similarities exist with rail applications, where efficient monitoring can be achieved by guided wave propagation along the rail. Dispersion curves and mode shapes are rather complicated due to the rail cross section (Hayashi et al. 2003; Bartoli et al. 2006), but dominant propagating modes with energy in the different sections of the rail can be identified and propagation along the rail achieved (Cawley et al. 2003). Monitoring systems to automatically detect rail breakage over long distances (km) have been developed (Loveday 2012). Accurate and repeatable experimental excitation and reception of specific guided wave modes is required to achieve good signal-to-noise ratio (SNR) for the detection of small defects.

---

## Experimental Methods

Guided wave measurements typically require different transducers and experimental equipment compared to standard bulk ultrasonic pulse-echo and pitch-catch measurements. As discussed briefly in section “[Guided Wave Propagation](#),” guided waves are multi-modal and dispersive, and measurements are often conducted at excitation frequencies in the kHz range below the cut-off frequencies of higher wave modes. Therefore, often the aim is to achieve selective experimental excitation of a specific guided wave mode with good control of the time signal and thus frequency content.

For higher frequency guided wave applications (typically in the lower MHz range), standard ultrasonic equipment and transducers can be employed (Fig. 12). However, in the kHz frequency range often employed in guided wave testing, the choice of transducers available from commercial manufacturers is more limited, in parts due to physical restrictions on the design of resonant piezoelectric transducers. For a number of applications, simple (and cheap) piezoelectric elements, dry-coupled, or glued to the specimen have been employed for guided wave measurements (Fig. 12). The element size and coupling influence the excited guided wave mode, as described, e.g., by Yu and Giurgiutiu (2012). Often measurements are conducted with a fixed excitation location using a permanently bonded piezoelectric element to allow comparison to



**Fig. 12** Picture of permanently bonded piezoelectric disc (with backing mass) and wedge transducer

theoretical configurations and to allow baseline subtraction. As the employed transducers are often non-resonant, the excitation time signal can be prescribed, e.g., as a narrow frequency bandwidth sinusoid (Fig. 4) by synthesizing the voltage signal using an arbitrary function generator. Using a voltage or power amplifier with the required frequency bandwidth, high excitation amplitudes with good signal-to-noise ratio (SNR) can be achieved. Typical approaches can be found in literature, e.g., Lowe et al. (1998), Giurgiutiu et al. (2002), Fromme et al. (2006).

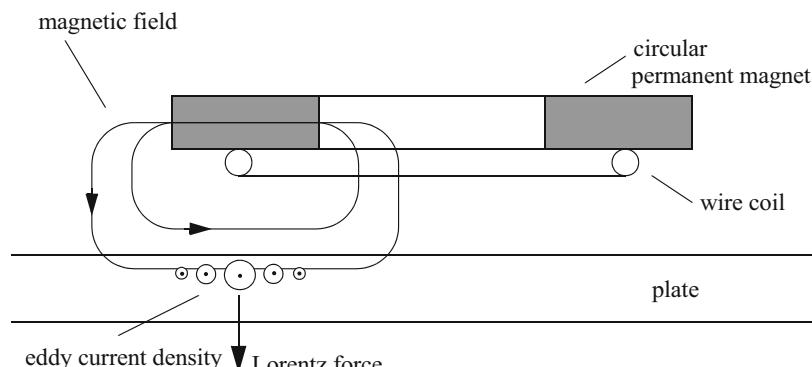
For phased or distributed arrays, piezoelectric elements have been used with good success to monitor large structures. However, for long-term monitoring, especially at higher temperatures and challenging environmental conditions, degradation of the bond layer used to couple the transducers and thus changes to the signals can occur (Attarian et al. 2014). For a number of applications, noncontact excitation and measurement methods have advantages, such as the lack of a coupling medium and possibility to scan the transducer along the structure.

Air-coupled transducers have been developed and are commercially available, either using a membrane or piezoelectric/static element. Controlling the angle and thus wavelength projection on the specimen, selective excitation of guided wave modes is possible (Castaings and Hosten 2008). Air-coupled excitation has a better efficiency for mode shapes with a significant out-of-plane component, but as the acoustic impedance of air and metals is significantly different, the overall excitation efficiency is limited.

Electromagnetic acoustic transducers (EMAT) allow for the non-contact excitation and reception of guided ultrasonic waves (Hirao and Ogi 2017). Typically, a static magnetic field is applied by a permanent magnet and an eddy current induced by an alternating current in a wire coil located close to the conductive specimen surface (Kawashima 1976). The interaction of the magnetic field and current generates a Lorentz force, which is perpendicular to both the magnetic field and eddy current. If the specimen material is ferromagnetic, in addition to the Lorentz force, magnetostrictive effects lead to induced stress and thus wave excitation of significantly increased amplitude (Jian et al. 2006). The usage of EMATs is described in guidelines such as the ASTM E1774–96 Standard Guide for Electromagnetic Acoustic Transducers.

EMATs do not require couplant, and thus are suitable for elevated temperatures, automated scanning, and guided wave excitation is not directly affected by surface roughness and coating. The wire coil to induce eddy currents can easily be designed to generate specific patterns, e.g., a meander or racetrack style coil to prescribe the wavelength of the excited wave mode. As the generated stresses are perpendicular to the magnetic field and eddy current (Fig. 13), EMATs are well-suited to excite guided wave modes with a significant in-plane component, e.g., shear horizontal (SH) waves. Limitations exist as the induced eddy currents depend on the stand-off distance and frequency. The skin depth of the eddy current density is inversely proportional to the frequency of the alternating current. For typical guided wave testing frequencies, the skin depth is less than 1 mm in steel. This limits the excitation amplitude for higher frequencies and thick specimens. A large current is required to obtain high amplitudes, requiring specialized electronic equipment such as power amplifiers and limiting the usage for explosion restricted applications. EMATs typically require strong magnets such as rare earth magnets, e.g., neodymium-boron, but could also employ electromagnets. The magnitude of the magnetostrictive effect depends on the specific material properties and can vary significantly, e.g., between different steel alloys. For sensors that target specifically the magnetostrictive effect, materials with controlled properties such as nickel or iron cobalt can be bonded to the structure, e.g., as a sleeve around pipes or with liquid couplant to allow rotation of a directional sensor (Vinogradov et al. 2018).

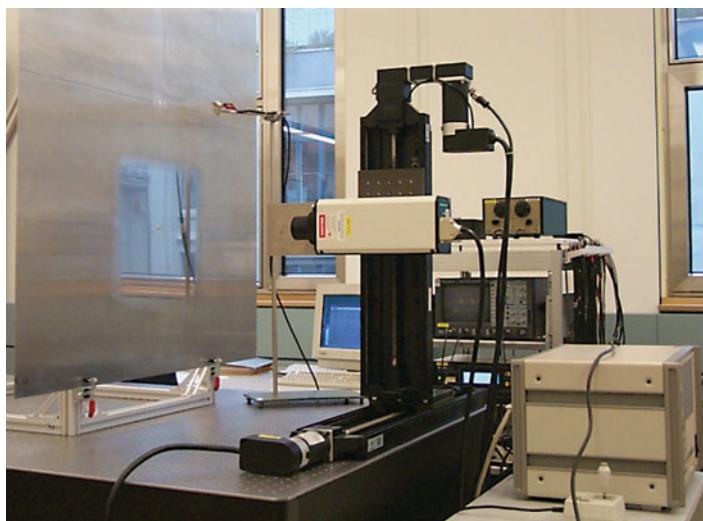
Lasers provide another non-contact excitation and measurement method widely used for laboratory studies (Zhang et al. 2006; Park et al. 2014), but due to safety constraints are more difficult to employ for field measurements. Typically, a pulsed laser beam is used to locally heat the material with the resulting thermal stresses exciting guided waves. Care must be taken to avoid too high laser intensity leading to material ablation. As most excitation lasers are class III or IV, the relevant safety precautions must be observed. The laser beam can be



**Fig. 13** Diagram of EMAT for excitation of  $A_0$  guided wave mode in plate with radially magnetized circular permanent magnet, wire coil, and Lorentz force due to induced eddy current shown schematically

optically shaped to achieve specific patterns and the pulse repetition rate can be matched to the required guided wave excitation frequency. However, as the excitation process relies on thermal processes with time delays mostly restricted to the specimen surface, exact control of the guided wave time signal and mode shape can be difficult to achieve.

Commercial laser interferometers or vibrometers (Fig. 14) exist that allow the pointwise measurement of guided waves with very good accuracy (Fromme and Sayir 2002b; Staszewski et al. 2007). The out-of-plane component of the specimen surface motion can be easily measured and quantified using either the phase interference to measure displacement or frequency shift to measure surface velocity. While instruments exist to measure in-plane motion directly, the repeatability and accuracy is typically lower than for out-of-plane measurement. As such, laser measurements of guided waves in plates have been used more widely for the  $A_0$  (flexural) mode with a significant out-of-plane (bending) motion, but the measurement of out-of-plane component of the  $S_0$  mode on untreated surfaces is also possible (Fromme et al. 2018). Using surface treatment (e.g., retro-reflective tape) and measurements with laser beams at three different angles, all three surface motion components can be quantified (Staszewski et al. 2007). Often either the laser interferometer or specimen are scanned relative to each other to measure guided wave propagation and scattering with good spatial accuracy to either quantify propagation speeds and thus material characteristics or detailed scattering patterns at defects. This methodology together with fixed piezoelectric excitation has been employed often, as it is well suited to the specific characteristics of guided waves and allows for comparison to predictions using numerical simulations.



**Fig. 14** Photograph of laser vibrometer mounted on scanning rig

## Simulations

For the understanding of guided wave propagation and scattering at defects, numerical simulations are useful to predict the behavior and understand phenomena better. Apart from theoretical calculations for simple geometries, different numerical methods with their respective advantages and drawbacks have been widely used. The most widely used method is Finite Element Analysis (FEA), as commercial software packages can be employed for computations. Other methods such as the Finite Difference Method (FDM), Boundary Elements, and peridynamic modeling have respective advantages, typically improved computational efficiency but often only limited standard, commercial software is available. The Semi-Analytical Finite Element (SAFE) method allows for the efficient computation of guided wave propagation along structures with irregular cross-sections and can be combined with local numerical approaches for the hybrid modeling of scattering, reflection, and transmission of guided waves.

A range of commercial FEA software packages, but also open-source programs (Huthwaite 2014), can be used to simulate guided wave propagation and scattering. Both explicit time marching and implicit schemes can be employed. Especially for implicit schemes, absorbing layers remove unwanted boundary reflections and allow smaller, more computationally efficient models to be used (Drozdz et al. 2006, Rajagopal et al. 2012, Shen and Giurgiutiu 2015). Stability criteria require small elements compared to the wavelength and short time steps for explicit time integration. For the long-propagation distances achievable with guided waves, this leads to large numbers of required elements and degrees of freedom. Depending on the NDE problem, 2D simulations of a cross-section of the structure can provide valuable insights with significant faster computation times. 2D FE models of wave propagation and scattering in composites to characterize impact damage have been developed (Pol and Banerjee 2013). The anisotropic material properties of composites can easily be implemented in commercial FE software and used in 3D FE models to study the scattering at delamination in composites (Murat et al. 2016; Ng et al. 2012). Different approaches for approximation of the damage using either Cartesian or adapted meshes can be used.

In order to efficiently predict guided wave propagation in complex wave guide geometries, e.g., stiffener bonded to a plate, the SAFE (Semi-Analytical Finite Element) method uses a 2D FE discretization of the cross-section of the waveguide, in combination with an analytical description of the behavior in the direction along the waveguide (Castaings and Lowe 2008). This significantly reduces the computational effort and has proven useful for irregularly shaped waveguides such as rails (Damljanovic and Weaver 2004). SAFE simulations can be conducted using flexible commercial or proprietary FE codes. Different implementations can be used, e.g., by Predoi et al. (2007), extended for complex propagating modes (Castaings and Lowe 2008) to allow for energy leakage. The dispersion diagram can be calculated by solving the eigenvalue solutions over the frequency range, to find the complex wavenumber, attenuation, and mode shape for each frequency.

SAFE methods for the guided wave propagation along waveguides can be combined with local FE or other numerical methods to calculate the scattering at

localized defects and the reflected and transmitted wave modes (Jezzine et al. 2018). Such hybrid models are computationally efficient but require care to achieve accurate coupling of the computational domains. Hybrid FE methods (Rose 2017) were used to study the scattering of guided waves at a hole (Paskaramoorthy et al. 1989) and to compare to experiments for the scattering of the fundamental  $S_0$  mode at a hole with a crack (Chang and Mal 1999). Hybrid boundary element methods were used to study Lamb wave reflection and mode conversion at a plate edge above the lowest cut-off frequency (Cho and Rose 1996). Hybrid FE modeling combined with the local interaction simulation approach (LISA) can be extended to allow for material anisotropy and damping (Shen and Cesnik 2016).

Finite Difference (FD) simulations based on a displacement formulation of the wave equations in isotropic, linear elastic media were used to study the scattering of guided waves (Harker 1984). Virieux (1986) presented a two-dimensional, stress-velocity FD formulation for modeling bulk wave propagation in heterogeneous media. The formulation was based on a system of first-order hyperbolic equations discretized on a staggered grid. This type of grid, first proposed by Madariaga (1976), has the useful property to minimize the number of variables per grid cell. A 3D displacement-velocity FD formulation was used to investigate the scattering of high frequency guided waves at cracks emanating from a fastener hole (Masserey and Fromme 2017). The Elastodynamic Finite Integration Technique (EFIT) (Fellinger et al. 1995) uses a similar formulation. The complex 3D geometry of impact damage in composites was implemented from X-ray computed tomography scans and the interaction of guided waves investigated (Leckey et al. 2014). For composites, the performance compared to different commercial FE solvers and experimental results were compared (Leckey et al. 2018). Other numerical methodologies such as the time domain spectral element method (Ostachowicz et al. 2012) and peri-ultrasound (Hafezi et al. 2017) have been employed to simulate guided wave propagation and scattering.

---

## Summary

The application of guided waves for nondestructive testing and monitoring has progressed significantly over the last decades. As guided waves are dispersive and multi-modal, a good understanding of the wave propagation and defect interaction in the structure of interest is required to ensure reliable defect detection and sensitivity. For the testing of pipelines, several commercially successful systems exist, allowing for the testing of pipe sections up to 100 m from a single access location and providing advantages over other NDE systems for a range of applications. However, the range of guided waves is more limited for coated structures and can be complicated for plate structures, requiring adaptation to the specific inspection problem and often numerical simulations to gain a full understanding of the wave physics. Specialized transducer and electronic equipment, often in the kHz frequency range, is required for the generation and reception of guided waves. Good control

over the guided wave mode and directionality helps to improve sensitivity for the detection of small defects. Further steps to gain widespread acceptance will require advanced simulations to quantify sensitivity and demonstrate applicability for complex industrial structures under real-life conditions.

---

## References

- Achenbach JD (1973) Wave propagation in elastic solids. Elsevier Science, Amsterdam
- Alleyne D, Cawley P (1991) A 2-dimensional Fourier-transform method for the measurement of propagating multimode signals. *J Acoust Soc Am* 89(3):1159–1168
- Alleyne D, Jones R, Vogt T (2017) GW test an introduction to long-range screening using guided waves. *Mater Eval* 75(10):1206–1213
- Attarian VA, Cegla FB, Cawley P (2014) Long-term stability of guided wave structural health monitoring using distributed adhesively bonded piezoelectric transducers. *Struct Health Monitor* 13(3):265–280
- Auld BA (1973) Acoustic fields and waves in solids. Hoboken. Wiley, New York
- Bai H, Shah AH, Popplewell N, Datta SK (2001) Scattering of guided waves by circumferential cracks in steel pipes. *J Appl Mech Trans ASME* 68(4):619–631
- Barshinger JN, Rose JL (2004) Guided wave propagation in an elastic hollow cylinder coated with a viscoelastic material. *IEEE Trans Ultrason Ferroelectr Freq Control* 51(11):1547–1556
- Bartoli I, Marzani A, Lanza di Scalea F, Viola E (2006) Modeling wave propagation in damped waveguides of arbitrary cross-section. *J Sound Vib* 295(3–5):685–707
- Brath AJ, Simonetti F, Nagy PB, Instanes G (2017) Guided wave tomography of pipe bends. *IEEE Trans Ultrason Ferroelectr Freq Control* 64(5):847–858
- Castaings M, Hosten B (2003) Guided waves propagating in sandwich structures made of anisotropic, viscoelastic, composite materials. *J Acoust Soc Am* 113(5):2622–2634
- Castaings M, Hosten B (2008) Ultrasonic guided waves for health monitoring of high-pressure composite tanks. *NDT & E Int* 41(8):648–655
- Castaings M, Lowe M (2008) Finite element model for waves guided along solid systems of arbitrary section coupled to infinite solid media. *J Acoust Soc Am* 123(2):696–708
- Castaings M, Singh D, Viot P (2012) Sizing of impact damages in composite materials using ultrasonic guided waves. *NDT & E Int* 46:22–31
- Cawley P, Lowe MJS, Alleyne DN, Pavlakovic B, Wilcox PD (2003) Practical long range guided wave testing: applications to pipes and rail. *Mater Eval* 61(1):66–74
- Chakrapani SK, Padhyar MJ, Balasubramaniam K (2012) Crack detection in full size Cz-silicon wafers using Lamb wave air coupled ultrasonic testing (LAC-UT). *J Nondestruct Eval* 31(1):46–55
- Chan H, Masserey B, Fromme P (2015) High frequency guided ultrasonic waves for hidden fatigue crack growth monitoring in multi-layer model aerospace structures. *Smart Mater Struct* 24(2), 025037
- Chang ZS, Mal A (1999) Scattering of Lamb waves from a rivet hole with edge cracks. *Mech Mater* 31(3):197–204
- Chapuis B, Terrien N, Royer D (2010) Excitation and focusing of Lamb waves in a multilayered anisotropic plate. *J Acoust Soc Am* 127(1):198–203
- Chen X, Michaels JE, Lee SJ, Michaels TE (2012) Load-differential imaging for detection and localization of fatigue cracks using Lamb waves. *NDT & E Int* 51:142–149
- Cho H, Lissenden CJ (2012) Structural health monitoring of fatigue crack growth in plate structures with ultrasonic guided waves. *Struct Health Monitor* 11(4):393–404
- Cho YH, Hongerholt DD, Rose JL (1997) Lamb wave scattering analysis for reflector characterization. *IEEE Trans Ultrason Ferroelectr Freq Control* 44(1):44–52

- Cho YH, Rose JL (1996) A boundary element solution for a mode conversion study on the edge reflection of Lamb waves. *J Acoust Soc Am* 99(4):2097–2109
- Croxford AJ, Moll J, Wilcox PD, Michaels JE (2010) Efficient temperature compensation strategies for guided wave structural health monitoring. *Ultrasonics* 50(4–5):517–528
- Dalton RP, Cawley P, Lowe MJS (2001) The potential of guided waves for monitoring large areas of metallic aircraft fuselage structure. *J Nondestruct Eval* 20(1):29–46
- Damljanovic V, Weaver RL (2004) Propagating and evanescent elastic waves in cylindrical waveguides of arbitrary cross section. *J Acoust Soc Am* 115(4):1572–1581
- Datta SK, Shah AH (2009) Elastic waves in composite media and structures with applications to ultrasonic nondestructive evaluation introduction. In: *Elastic waves in composite media and structures: with applications to ultrasonic nondestructive evaluation*. CRC Press, Boca Raton, pp 1–9
- Davies J, Cawley P (2009) The application of synthetic focusing for imaging crack-like defects in pipelines using guided waves. *IEEE Trans Ultrason Ferroelectr Freq Control* 56(4):759–771
- Demma A, Cawley P, Lowe M (2003) Scattering of the fundamental shear horizontal mode from steps and notches in plates. *J Acoust Soc Am* 113(4):1880–1891
- Di Scalea FL, Matt H, Bartoli I, Coccia S, Park G, Farrar C (2007) Health monitoring of UAV wing skin-to-spar joints using guided waves and macro fiber composite transducers. *J Intell Mater Syst Struct* 18(4):373–388
- Diamanti K, Hodgkinson JM, Soutis C (2004) Detection of low-velocity impact damage in composite plates using Lamb waves. *Struct Health Monitor* 3(1):33–41
- Diligent O, Grahn T, Bostrom A, Cawley P, Lowe MJS (2002) The low-frequency reflection and scattering of the S-0 Lamb mode from a circular through-thickness hole in a plate: finite element, analytical and experimental studies. *J Acoust Soc Am* 112(6):2589–2601
- Dixon S, Jaques D, Palmer SB, Rowlands G (2004) The measurement of shear and compression waves in curing epoxy adhesives using ultrasonic reflection and transmission techniques simultaneously. *Meas Sci Technol* 15(5):939–947
- Doherty C, Chiu WK (2012) Scattering of ultrasonic-guided waves for health monitoring of fuel weep holes. *Struct Health Monitor* 11(1):27–42
- Drozdz M, Moreau L, Castaings M, Lowe MJS, Cawley P (2006). Efficient numerical modelling of absorbing regions for boundaries of guided waves problems. In: Thompson DO, Chimenti DE (eds) *Review of progress in quantitative nondestructive evaluation*, vols 25a and 25b, vol 820. American Institute of Physics, Melville, pp 126–133
- Fan Z, Castaings M, Lowe MJS, Bateau C, Fromme P (2013) Feature-guided waves for monitoring adhesive shear modulus in bonded stiffeners. *NDT & E Int* 54:96–102
- Fan Z, Lowe MJS (2009) Elastic waves guided by a welded joint in a plate. *Proc R Soc Math Phys Eng Sci* 465(2107):2053–2068
- Fellinger P, Marklein R, Langenberg KJ, Klaholz S (1995) Numerical modeling of elastic-wave propagation and scattering with efit - elastodynamic finite integration technique. *Wave Motion* 21(1):47–66
- Flynn EB, Todd MD, Wilcox PD, Drinkwater BW, Croxford AJ (2011) Maximum-likelihood estimation of damage location in guided-wave structural health monitoring. *Proc R Soc Math Phys Eng Sci* 467(2133):2575–2596
- Freemantle RJ, Challis RE (1998) Combined compression and shear wave ultrasonic measurements on curing adhesive. *Meas Sci Technol* 9(8):1291–1302
- Fromme P, Pizzolato M, Robyr JL, Masserey B (2018) Lamb wave propagation in monocrystalline silicon wafers. *J Acoust Soc Am* 143(1):287–295
- Fromme P, Rouge C (2011) Directivity of guided ultrasonic wave scattering at notches and cracks C3. *J Phys Conf Ser* 269:1
- Fromme P, Sayir MB (2002a) Detection of cracks at rivet holes using guided waves. *Ultrasonics* 40(1–8):199–203
- Fromme P, Sayir MB (2002b) Measurement of the scattering of a Lamb wave by a through hole in a plate. *J Acoust Soc Am* 111(3):1165–1170

- Fromme P, Wilcox PD, Lowe MJS, Cawley P (2006) On the development and testing of a guided ultrasonic wave array for structural integrity monitoring. *IEEE Trans Ultrason Ferroelectr Freq Control* 53(4):777–784
- Giuriutiu V, Zagrai A, Bao JJ (2002) Piezoelectric wafer embedded active sensors for aging aircraft structural health monitoring. *Struct Health Monitor* 1(1):41–61
- Graff KF (1975) Wave motion in elastic solids. Oxford University Press, New York
- Greve DW, Zheng P, Oppenheim IJ (2008) The transition from Lamb waves to longitudinal waves in plates. *Smart Mater Struct* 17(3), 035029
- Gridin D, Craster RV, Fong J, Lowe MJS, Beard M (2003) The high-frequency asymptotic analysis of guided waves in a circular elastic annulus. *Wave Motion* 38(1):67–90
- Grondel S, Paget C, Delebarre C, Assaad J, Levin K (2002) Design of optimal configuration for generating a(0) Lamb mode in a composite plate using piezoceramic transducers. *J Acoust Soc Am* 112(1):84–90
- Guo N, Cawley P (1993) The interaction of Lamb waves with delaminations in composite laminates. *J Acoust Soc Am* 94(4):2240–2246
- Guy P, Jayet Y, Goujon L (2003) Guided waves interaction with complex delaminations. Application to damage detection in composite structures. In: Kundu T (ed) Smart nondestructive evaluation and health monitoring of structural and biological systems II. SPIE 5047, pp 25–33
- Hafezi MH, Alebrahim R, Kundu T (2017) Peri-ultrasound for modeling linear and nonlinear ultrasonic response. *Ultrasonics* 80:47–57
- Hall JS, Fromme P, Michaels JE (2014) Guided wave damage characterization via minimum variance imaging with a distributed Array of ultrasonic sensors. *J Nondestruct Eval* 33(3):299–308
- Harker AH (1984) Numerical modelling of the scattering of elastic waves in plates. *J Nondestruct Eval* 4(2):89–106
- Hayashi T, Kawashima K, Sun ZQ, Rose JL (2005) Guided wave focusing mechanics in pipe. *J Pressure Vessel Technol-Trans ASME* 127(3):317–321
- Hayashi T, Song WJ, Rose JL (2003) Guided wave dispersion curves for a bar with an arbitrary cross-section, a rod and rail example. *Ultrasonics* 41(3):175–183
- Herdovics B, Cegla F (2018) Structural health monitoring using torsional guided wave electromagnetic acoustic transducers. *Struct Health Monitor* 17(1):24–38
- Hirao M, Ogi H (2017) Brief instruction to build EMATs. Electromagnetic acoustic transducers: noncontacting ultrasonic measurements using Emats, 2nd edn. Springer, Tokyo, pp 69–79
- Hosten B, Castaings M (1993) Transfer-matrix of multilayered absorbing and anisotropic media - measurements and simulations of ultrasonic wave-propagation through composite-materials. *J Acoust Soc Am* 94(3):1488–1495
- Howard R, Cegla F (2017) Detectability of corrosion damage with circumferential guided waves in reflection and transmission. *NDT & E Int* 91:108–119
- Huthwaite P (2014) Accelerated finite element elastodynamic simulations using the GPU. *J Comput Phys* 257:687–707
- Jezzine K, Imperiale A, Demaldent E, Le Bourdais F, Calmon P, Dominguez N (2018) Modeling approaches for the simulation of ultrasonic inspections of anisotropic composite structures in the CIVA software platform. In: Chimenti DE, Bond LJ (eds) 44th annual review of progress in quantitative nondestructive evaluation, vol 37, American Institute of Physics, Melville, 1949
- Jian X, Dixon S, Edwards RS, Morrison J (2006) Coupling mechanism of an EMAT. *Ultrasonics* 44:E653–E656
- Kawashima K (1976) Experiments with 2 types of electromagnetic ultrasonic transducers. *J Acoust Soc Am* 60(2):365–373
- Kazys R, Demcenko A, Zukauskas E, Mazeika L (2006) Air-coupled ultrasonic investigation of multi-layered composite materials. *Ultrasonics* 44:E819–E822
- Khalili P, Cawley P (2016) Excitation of single-mode Lamb waves at high-frequency-thickness products. *IEEE Trans Ultrason Ferroelectr Freq Control* 63(2):303–312

- Konstantinidis G, Drinkwater BW, Wilcox PD (2006) The temperature stability of guided wave structural health monitoring systems. *Smart Mater Struct* 15(4):967–976
- Kostson E, Fromme P (2009) Fatigue crack growth monitoring in multi-layered structures using guided ultrasonic waves. *J Phys Conf Ser* 195
- Kundu T, Das S, Martin SA, Jata KV (2008) Locating point of impact in anisotropic fiber reinforced composite plates. *Ultrasonics* 48(3):193–201
- Le Crom B, Castaings M (2010) Shear horizontal guided wave modes to infer the shear stiffness of adhesive bond layers. *J Acoust Soc Am* 127(4):2220–2230
- Leckey CAC, Rogge MD, Parker FR (2014) Guided waves in anisotropic and quasi-isotropic aerospace composites: three-dimensional simulation and experiment. *Ultrasonics* 54(1):385–394
- Leckey CAC, Wheeler KR, Hafichuk VN, Hafichuk H, Timucin DA (2018) Simulation of guided-wave ultrasound propagation in composite laminates: benchmark comparisons of numerical codes and experiment. *Ultrasonics* 84:187–200
- Lee BC, Staszewski WJ (2007) Lamb wave propagation modelling for damage detection: II. Damage monitoring strategy. *Smart Mater Struct* 16(2):260–274
- Leinov E, Lowe MJS, Cawley P (2015) Investigation of guided wave propagation and attenuation in pipe buried in sand. *J Sound Vib* 347:96–114
- Leleux A, Micheau P, Castaings M (2013) Long range detection of defects in composite plates using Lamb waves generated and detected by ultrasonic phased array probes. *J Nondestruct Eval* 32(2):200–214
- Leonard KR, Malyarenko EV, Hinders MK (2002) Ultrasonic Lamb wave tomography. *Inverse Problem* 18(6):1795–1808
- Levine RM, Michaels JE (2013) Model-based imaging of damage with Lamb waves via sparse reconstruction. *J Acoust Soc Am* 133(3):1525–1534
- Li J, Rose JL (2001) Implementing guided wave mode control by use of a phased transducer array. *IEEE Trans Ultrason Ferroelectr Freq Control* 48(3):761–768
- Lindgren E, Aldrin JC, Jata K, Scholes B, Knopp J (2007) Ultrasonic plate waves for fatigue crack detection in multi-layered metallic structures. In: Kundu T (eds) *Health monitoring of structural and biological systems 2007*. Proceedings of SPIE 6532
- Liu GL, Qu JM (1998) Guided circumferential waves in a circular annulus. *J Appl Mech-Trans ASME* 65(2):424–430
- Loveday PW (2012) Guided wave inspection and monitoring of railway track. *J Nondestruct Eval* 31(4):303–309
- Lovstad A, Cawley P (2012) The reflection of the fundamental torsional mode from pit clusters in pipes. *NDT & E Int* 46:83–93
- Lowe MJS, Alleyne DN, Cawley P (1998) The mode conversion of a guided wave by a part-circumferential notch in a pipe. *J Appl Mech-Trans ASME* 65(3):649–656
- Lowe MJS, Cawley P (1994) The applicability of plate wave techniques for the inspection of adhesive and diffusion bonded joints. *J Nondestruct Eval* 13(4):185–200
- Lowe MJS, Diligent O (2002) Low-frequency reflection characteristics of the s(0) Lamb wave from a rectangular notch in a plate. *J Acoust Soc Am* 111(1):64–74
- Lu Y, Michaels JE (2009) Feature extraction and sensor fusion for ultrasonic structural health monitoring under changing environmental conditions. *IEEE Sensors J* 9(11):1462–1471
- Madariaga R (1976) Dynamics of an expanding circular fault. *Bull Seismol Soc Am* 66(3):639–666
- Mal AK, Xu PC, Bar Cohen Y (1989) Analysis of leaky Lamb waves in bonded plates. *Int J Eng Sci* 27(7):779–791
- Malyarenko EV, Hinders MK (2000) Fan beam and double crosshole Lamb wave tomography for mapping flaws in aging aircraft structures. *J Acoust Soc Am* 108(4):1631–1639
- Masserey B, Fromme P (2017) Analysis of high frequency guided wave scattering at a fastener hole with a view to fatigue crack detection. *Ultrasonics* 76:78–86
- McKeon JCP, Hinders MK (1999) Lamb wave scattering from a through hole. *J Sound Vib* 224(5):843–862

- Mesnil O, Leckey CAC, Ruzzene M (2014) Instantaneous wavenumber estimation for damage quantification in layered plate structures. In: Kundu T (ed) Health monitoring of structural and biological systems 2014. Proceedings of SPIE 9064
- Michaels JE (2008) Detection, localization and characterization of damage in plates with an in situ array of spatially distributed ultrasonic sensors. *Smart Mater Struct* 17(3), 17 035035
- Michaels JE, Lee SJ, Croxford AJ, Wilcox PD (2013) Chirp excitation of ultrasonic guided waves. *Ultrasonics* 53(1):265–270
- Mudge PJ (2001) Field application of the Teletest (R) long-range ultrasonic testing technique. *Insight* 43(2):74–77
- Murat BIS, Khalili P, Fromme P (2016) Scattering of guided waves at delaminations in composite plates. *J Acoust Soc Am* 139(6):3044–3052
- Nagy PB, Adler L (1989) Nondestructive evaluation of adhesive joints by guided-waves. *J Appl Phys* 66(10):4658–4663
- Ng C-T, Veidt M (2011) Scattering of the fundamental anti-symmetric Lamb wave at delaminations in composite laminates. *J Acoust Soc Am* 129(3):1288–1296
- Ng CT, Veidt M, Rose LRF, Wang CH (2012) Analytical and finite element prediction of Lamb wave scattering at delaminations in quasi-isotropic composite laminates. *J Sound Vib* 331(22):4870–4883
- Ostachowicz W, Kudela P, Krawczuk M, Zak A (2012) Guided waves in structures for SHM: the time-domain spectral element method. Blackwell Science Publ, Oxford
- Pao YH, Chao CC (1964) Diffractions of flexural waves by a cavity in an elastic plate. *AIAA J* 2(11):2004–2010
- Park B, An YK, Sohn H (2014) Visualization of hidden delamination and debonding in composites through noncontact laser ultrasonic scanning. *Compos Sci Technol* 100:10–18
- Paskaramoorthy R, Shah AH, Datta SK (1989) Scattering of flexural waves by cavities in a plate. *Int J Solids Struct* 25(10):1177–1191
- Pavlakovic, B., M. Lowe, D. Alleyne and P. Cawley (1997). “DISPERSE: A general purpose program for creating dispersion curves,” in Review of Progress in Quantitative NDE, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York), Vol. 16, pp. 185–192
- Pol CB, Banerjee S (2013) Modeling and analysis of propagating guided wave modes in a laminated composite plate subject to transient surface excitations. *Wave Motion* 50(5):964–978
- Potel C, Baly S, de Belleval JF, Lowe M, Gatignol P (2005) Deviation of a monochromatic Lamb wave beam in anisotropic multilayered media: asymptotic analysis, numerical and experimental results. *IEEE Trans Ultrason Ferroelectr Freq Control* 52(6):987–1001
- Prada C, Clorennec D, Murray TW, Royer D (2009) Influence of the anisotropy on zero-group velocity Lamb modes. *J Acoust Soc Am* 126(2):620–625
- Predoi MV, Castaings M, Hosten B, Bacon C (2007) Wave propagation along transversely periodic structures. *J Acoust Soc Am* 121(4):1935–1944
- Puthillath P, Rose JL (2010) Ultrasonic guided wave inspection of a titanium repair patch bonded to an aluminum aircraft skin. *Int J Adhes Adhes* 30(7):566–573
- Quarry MJ (2004) Guided wave inspection of multi-layered structures. In: Quantitative nondestructive evaluation. AIP conference proceedings, vol 700, pp 246–253
- Rajagopal P, Drozdz M, Skelton EA, Lowe MJS, Craster RV (2012) On the use of absorbing layers to simulate the propagation of elastic waves in unbounded isotropic media using commercially available finite element packages. *NDT & E Int* 51:30–40
- Rakow A, Chang F-K (2012) A structural health monitoring fastener for tracking fatigue crack growth in bolted metallic joints. *Struct Health Monitor* 11(3):253–267
- Ramadas C, Balasubramaniam K, Joshi M, Krishnamurthy CV (2010) Interaction of guided Lamb waves with an asymmetrically located delamination in a laminated composite plate. *Smart Mater Struct* 19(6), 065009
- Ratassep M, Fletcher S, Lowe MJS (2010) Scattering of the fundamental torsional mode at an axial crack in a pipe. *J Acoust Soc Am* 127(2):730–740

- Ratnam D, Balasubramaniam K, Maxfield BW (2012) Generation and detection of higher-order mode clusters of guided waves (HOMC-GW) using meander-coil EMATs. *IEEE Trans Ultrason Ferroelectr Freq Control* 59(4):727–737
- Rokhlin SI, Hefets M, Rosen M (1981) An ultrasonic interface-wave method for predicting the strength of adhesive bonds. *J Appl Phys* 52(4):2847–2851
- Rose JL (2002a) A baseline and vision of ultrasonic guided wave inspection potential. *J Pres Vessel Technol Tran ASME* 124(3):273–282
- Rose JL (2002b) Standing on the shoulders of giants: an example of guided wave inspection. *Mater Eval* 60(1):53–59
- Rose JL (2014) Ultrasonic guided waves in solid media. Cambridge University Press, Cambridge
- Rose JL (2017) Aspects of a hybrid analytical finite element method approach for ultrasonic guided wave inspection design. *J Nondestruct Eval Diagnos Prognos Eng Syst* 1(1):011001
- Rose JL, Zhang L, Avioli MJ, Mudge PJ (2005) A natural focusing low frequency guided wave experiment for the detection of defects beyond elbows. *J Press Vessel Technol-Trans ASME* 127(3):310–316
- Salas KI, Cesnik CES (2009) Guided wave excitation by a CLoVER transducer for structural health monitoring: theory and experiments. *Smart Mater Struct* 18(7):1–27
- Sanderson RM, Hutchins DA, Billson DR, Mudge PJ (2013) The investigation of guided wave propagation around a pipe bend using an analytical modeling approach. *J Acoust Soc Am* 133(3):1404–1414
- Sargent JP (2006) Corrosion detection in welds and heat-affected zones using ultrasonic Lamb waves. *Insight* 48(3):160–167
- Seifried R, Jacobs LJ, Qu JM (2002) Propagation of guided waves in adhesive bonded components. *NDT & E Int* 35(5):317–328
- Shen YF, Cesnik CES (2016) Hybrid local FEM/global LISA modeling of damped guided wave propagation in complex composite structures. *Smart Mater Struct* 25(9):20
- Shen YF, Giurgiutiu V (2015) Effective non-reflective boundary for Lamb waves: Theory, finite element implementation, and applications. *Wave Motion* 58:22–41
- Staszewski WJ, Lee BC, Traynor R (2007) Fatigue crack detection in metallic structures with Lamb waves and 3D laser vibrometry. *Meas Sci Technol* 18(3):727–739
- Su Z, Ye L, Lu Y (2006) Guided Lamb waves for identification of damage in composite structures: a review. *J Sound Vib* 295(3–5):753–780
- Terrien N, Osmont D, Royer D, Lepoutre F, Deom A (2007) A combined finite element and modal decomposition method to study the interaction of Lamb modes with micro-defects. *Ultrasonics* 46(1):74–88
- Thompson RB, Thompson DO (1991) Past experiences in the development of tests for adhesive bond strength. *J Adhes Sci Technol* 5(8):583–599
- Van Velsor JK, Rose JL, Nestleroth JB (2009) Enhanced coating disbond detection capabilities in pipe using circumferential shear horizontal guided waves. *Mater Eval* 67(10):1179–1188
- Veidt M, Sachse W (1994) Ultrasonic point-source point-receiver measurements in thin specimens. *J Acoust Soc Am* 96(4):2318–2326
- Velichko A, Wilcox PD (2008) Guided wave arrays for high resolution inspection. *J Acoust Soc Am* 123(1):186–196
- Viktorov IA (1967) Rayleigh and Lamb waves - physical theory and applications. Plenum, New York
- Vinogradov S, Eason T, Lozev M (2018) Evaluation of magnetostrictive transducers for guided wave monitoring of pressurized pipe at 200 degrees C. *J Pres Vessel Technol-Tran ASME* 140(2):7
- Virieux J (1986) P-SV-wave propagation in heterogeneous media – velocity-stress finite-difference method. *Geophysics* 51(4):889–901
- Wang CH, Rose JT, Chang FK (2004) A synthetic time-reversal imaging method for structural health monitoring. *Smart Mater Struct* 13(2):415–423

- Wilcox PD (2003) A rapid signal processing technique to remove the effect of dispersion from guided wave signals. *IEEE Trans Ultrason Ferroelectr Freq Control* 50(4):419–427
- Wilcox PD, Lowe M, Cawley P (2005) Omnidirectional guided wave inspection of large metallic plate structures using an EMAT array. *IEEE Trans Ultrason Ferroelectr Freq Control* 52(4):653–665
- Yu L, Giurgiutiu V (2012) Piezoelectric wafer active sensors in Lamb wave-based structural health monitoring. *JOM* 64(7):814–822
- Zhang F, Krishnaswamy S, Lilley CM (2006) Bulk-wave and guided-wave photoacoustic evaluation of the mechanical properties of aluminum/silicon nitride double-layer thin films. *Ultrasoundics* 45(1–4):66–76
- Zhao X, Royer RL, Owens SE, Rose JL (2011) Ultrasonic Lamb wave tomography in structural health monitoring. *Smart Mater Struct* 20(10):10
- Zhao XL, Rose JL (2004a) Guided circumferential shear horizontal waves in an isotropic hollow cylinder. *J Acoust Soc Am* 115(5):1912–1916
- Zhao, X. L. and J. L. Rose (2004b). Three-dimensional defect in a plate boundary element modeling for guided wave scattering. In: Lee SS, Yoon DJ, Lee JH, Lee S (eds) *Advances in nondestructive evaluation*, Pt 1–3. Trans Tech Publications, pp 270–273: 453–460



# Laser-Induced Surface Acoustic Waves for Material Testing

6

Dieter Schneider

## Contents

Introduction .....	172
Surface Acoustic Waves .....	175
Laser Generation of Surface Acoustic Waves .....	184
Experimental Setup and Measuring Procedure .....	188
Effect of Film and Substrate on the Surface Wave Dispersion .....	194
The Effect of the Film Material .....	194
The Effect of the Substrate Material .....	195
The Effect of Film Thickness .....	196
The Effect of a Double Layer of Two Films with Different Properties .....	199
Applications .....	201
Super-Hard Tetrahedral Amorphous Carbon (ta-C) .....	201
Porous Titanium Films .....	204
Thermal-Sprayed Coatings .....	207
Subsurface Damage .....	213
Laser-Harden Steels .....	219
Comparing the Results with Other Methods .....	225
Summary .....	228
References .....	229

## Abstract

Surface acoustic waves are elastic vibrations which propagate along the surface of the material. They are very sensitive to films and surface treatments, since the wave energy is concentrated near the surface. Therefore, there is a growing interest in using this acoustic wave mode for nondestructive testing. Whereas the wave velocity is constant for homogenous materials, the velocity  $c$  depends on frequency  $f$  for coated and surface-modified materials. This phenomenon, termed dispersion,

---

D. Schneider (✉)

Fraunhofer Institute for Material and Beam Technology (IWS), Dresden, Germany  
e-mail: [dieter.schneider@iws.fraunhofer.de](mailto:dieter.schneider@iws.fraunhofer.de)

can be used to determine important film parameters such as Young's modulus, density, or film thickness. Especially, Young's modulus is an interesting parameter for nondestructive characterization of film materials, since it depends on the bonding conditions and the microstructure. In order to determine the parameters of the film material, the dispersion curve  $c(f)$  is measured and fitted by a theoretical curve. Many experimental setups use pulse lasers to create surface acoustic waves. Short laser pulses can create wideband acoustic impulses. The laser is a non-contact acoustic source that can precisely be positioned on the material surface, which enables an accurate measurement of the dispersion. Five examples of application are presented which demonstrate that surface acoustic waves can be used for very different problems of surface characterization: diamond-like carbon films (ta-C) with thickness down to few nanometers, porous metal films of titanium with a thickness in the micrometer range, thermal-sprayed ceramic coatings with a thickness of some hundreds of micrometers, laser-hardened steels up to the depth of one millimeter, and subsurface damage in semiconductor materials.

---

## Introduction

The surface is an essential part of technical components. Engineering components suffer destructive effects on their surfaces, like wear, fatigue, or corrosion. Protecting the surface by hardening treatments or coatings is a widespread way to extend the life and improve the quality of the products. In the semiconductor industry, thin film technology is the basis of the functionality of microelectronic circuits. For these complex structures of micro- and nano-products, optimizing and matching the mechanical material properties are also very important due to residual stresses and thermal fatigue that influence their reliability.

Important properties of surface films and coatings are the thickness, the adhesion to the substrate, the hardness, the elastic modulus, thermal conductivity, or electrical conductivity. Sophisticated techniques are available for chemical surface analyses with a resolution up to the range of atomic layers, such as Auger electron spectroscopy, secondary ion mass spectroscopy, scanning tunneling microscopy, and electron microscopy (O'Connor et al. 1992).

Apart from chemical and topographic analyses, the interest is also focused on the mechanical characterization of material surfaces. Important mechanical test methods are the instrumented indentation test (Oliver and Pharr 1992), the bending test, tensile bond test, scratch test, and cavitation test (Mittal 1976; Ollendorf and Schneider 1999).

In the last two decades, surface acoustic waves have increasingly proven to be a physical test method which is a suitable analyzing tool that not only complements the traditional mechanical test methods but can also provide insight into new material phenomena.

Two fields of application can be distinguished, the surface flaw detection and surface material characterization. The surface material characterization refers to the

measurement of material parameters Young's modulus, density and thickness of surface films and near-surface regions of the material. Both fields of application use the high sensitivity of the surface acoustic wave to material surface modifications, since the wave amplitude is highest at the surface and decays within the material exponentially. The sensing depth of the wave can be varied by varying the frequency of the wave,  $f$ . The penetration depth of the wave is defined to be equal to the wavelength  $\lambda$ . The penetration depth is lower, the higher the frequency  $f$  due to the relation  $\lambda=c/f$  ( $c$ : phase velocity). Films with thickness of a few nanometers on substrates with low ultrasonic damping such as silicon single crystals can be investigated, using high-frequency waves ( $f > 200$  MHz), just as laser-hardened zones with a depth of more than 1 mm can be analyzed using frequencies in the range of  $f < 5$  MHz.

It is obvious to apply spectroscopic methods to obtain information about the microstructure from different depths of the material. Pulsed lasers are suitable instruments to generate acoustic wave pulses. Laser pulses with duration in the range of few nanoseconds can generate acoustic pulses with a bandwidth of some hundreds of megacycles. Pulse lasers are still expensive sound sources compared to piezoelectric acoustic sensors. However, for some applications, lasers are indispensable sources of ultrasound, since they can be positioned with high precision over large dimensions. They are contactless and very reproducible, which avoids the problem of reproducible coupling of the sensor to the surface of the test sample. Lasers can also be used at high temperatures. For nondestructive testing, the laser sound source is usually applied in the thermoelastic regime to avoid damage of the surface by material ablation.

Laser-acoustic devices use both contact and contactless methods for the detection of the acoustic waveform. In the first case, an acoustic sensor typically based on the piezoelectric effect is placed on the surface. This is a low-cost method, which has the disadvantage that usually a liquid coupling medium is needed between the sensor and sample surface, and the frequency range may be limited to that of commercially available acoustic sensors. For surface acoustic waves, Coufal et al. (1992) have proposed a simple PVDF-foil sensor which can be placed directly on the material surface without a coupling medium and enables surface acoustic waves in the frequency range beyond 200 MHz to be detected.

There are applications for which contact to the surface of the material has to be avoided, for example in the semiconductor industry. This requires that the detection of the acoustic signal must also be contactless. The technique of laser-interferometry is frequently used for these applications.

In this contribution, surface acoustic waves are used for nondestructive characterization of material surfaces modified by films or surface treatments. The idea is to measure the propagation velocity, which depends on the elastic parameters and the density of the material. For a homogeneous and isotropic material described by Young's modulus  $E$ , Poisson's ratio  $\nu$  and density  $\rho$ , Eq. 1 gives a good approximation for the phase velocity  $c$  of the surface acoustic wave (Farnell 1970):

$$c = \frac{0.87 + 1.12\nu}{1 + \nu} \sqrt{\frac{E}{2\rho(1 + \nu)}} \quad (1)$$

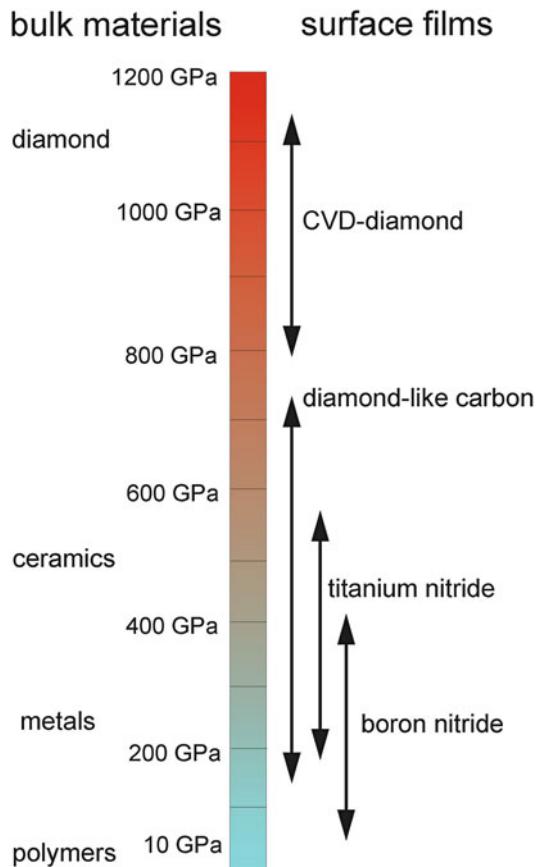
Young's modulus is a fundamental mechanical parameter, which is known from Hook's law. The elastic behavior of isotropic media is completely described by Young's modulus  $E$  and Poisson's ratio  $\nu$ . For anisotropic materials, such as single crystals, the elastic constants  $C_{ij}$  are used. Measuring the elastic modulus of thin films and coatings becomes increasingly important, for two reasons. First, it is a main mechanical parameter that characterizes the elastic stiffness of the material. It determines the residual stresses as well the elastic energy induced by external loading, which is responsible for the initiation of cracks and micro-fractures. Second, for important covalent and ionic bonded materials Young's modulus correlates with the hardness which is difficult to measure for thin films. Young's modulus is intimately related to the atomic structure, which can vary within a wide range for thin film materials, depending on the deposition conditions. Thin films often show metastable, amorphous, or defect-rich structures. Tabulated values of the elastic modulus measured for bulk materials may deviate considerably from the modulus of the same material deposited as thin films. For many film materials, the bulk properties are not known since they cannot be made as bulk sample.

Figure 1 shows the range of Young's modulus for some bulk materials and materials deposited as films. Diamond has the highest value, 1142 GPa, ceramic materials up to 500 GPa, metals up to 300 GPa, and polymers in the range of 10 GPa. The values given for some film materials show that their Young's modulus can vary within a wide range depending on the deposition technology (Kuschnereit et al. 1995; Carvalho et al. 2001; Leonhardt et al. 2004; Rebholz et al. 2006; Hess 2009). Therefore, Young's modulus is a sensitive parameter for the film quality. For example, Young's modulus of tetrahedral amorphous diamond-like carbon ta-C correlates with the ratio of  $sp^3/sp^2$ -bonds. Measuring Young's modulus enables one to conclude on the diamond-likeness of these films (Schultrich et al. 1996).

In the case that the surface acoustic wave propagates in a material which is coated with a surface film, the wave velocity  $c$  depends on the ratio of film thickness  $d$  to the wavelength  $\lambda$ . Even for the simple case of a homogenous isotropic film on a homogenous isotropic substrate, a simple relation similar to Eq. 1 cannot be given for the phase velocity  $c$ , depending on the material parameters. The phase velocity has to be found from a numerical solution of a system of equations which is deduced from the boundary conditions of stress and displacements at the surface and at the interface between film and substrate. The dispersion relation  $c = \omega/k$  ( $\omega = 2\pi f$ : circular frequency,  $k = 2\pi/\lambda$ : wave vector) is equal to the roots of the characteristic polynomial of this system of equations.

In this contribution, the properties of surface acoustic waves are described. Experimental setups are presented which measure the dispersion of surface acoustic waves in a wide frequency range. The useful application of surface wave method is demonstrated for some very different examples: testing of super-hard diamond-like carbon films, evaluation of porosity in metal films made of titanium, evaluation of

**Fig. 1** The range of Young's modulus for some bulk materials and materials deposited as thin films on a substrate (Schneider et al. 2012a)



the defective microstructure in thermal-sprayed coatings, determination of the depth of subsurface damage in semiconductor materials, and the characterization of laser-hardened steel. Finally, the results obtained with the surface acoustic waves are compared with the alternative method of instrumented indentation test.

## Surface Acoustic Waves

In an infinite homogenous medium with isotropic linear-elastic properties, the equation of elastic wave motion has two independent solutions, longitudinal and transverse (shear) waves. For longitudinal waves, the material particles vibrate in the direction of wave propagation and for transverse waves perpendicular to the direction of wave propagation. Lord Rayleigh (1885) proposed that at the free surface of an elastic half-space a surface wave mode can propagate, which has to be a superposition of both, longitudinal and transverse vibration. If it is required, that the

wave energy does not leak into the bulk of the material, the wave amplitude has to decay exponentially from the surface into the material. Therefore, the wave energy is concentrated at the surface, which is the reason that surface acoustic waves are very sensitive to surface layers and surface modifications.

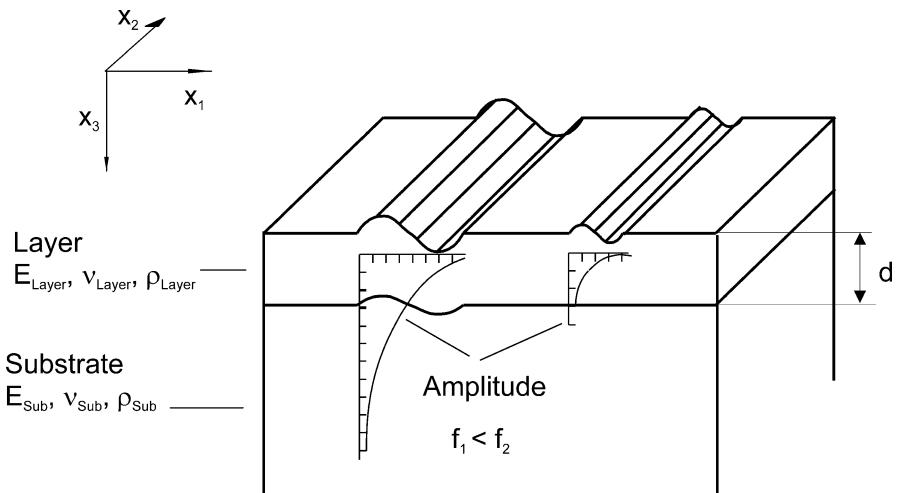
Surface acoustic waves are not plane waves in the original sense, because their amplitude varies with depth. Therefore, they are also termed “straight-crested” waves, considering that there is no variation of the displacement parallel to the surface and perpendicular to the direction of propagation (Farnell 1970). The particle motion at the surface and in the depth is elliptical. The velocity of the surface acoustic waves is lower than the velocity of the shear wave.

A coordinate system with axes  $x_1$ ,  $x_2$ , and  $x_3$  is defined (Fig. 2). The surface of the material is the  $x_1-x_2$ -plane and the surface acoustic wave propagates along the  $x_1$ -direction. The wave motion takes place in the  $x_1-x_3$ -plane, also termed the sagittal plane.  $x_3$  is positive into the material. For a single crystal, the sagittal plane should be a plane of crystallographic symmetry.

The equation of elastic wave motion is deduced from the condition that the force of inertia to accelerate a particle in the material in the coordinate direction  $i$  is equal to the sum of the derivatives of the stresses  $\sigma_{ik}$  into this direction (Landau and Lifschitz 1975).

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \frac{\partial \sigma_{ik}}{\partial x_k} \quad (i, k = 1, 2, 3) \quad (2)$$

For isotropic elastic media, the equation of elastic wave motion has the following form (Landau and Lifschitz 1975):



**Fig. 2** Amplitude of a surface acoustic wave propagating in coated material, the influence of frequency  $f$  on the penetration depth, Young's modulus  $E$ , Poisson's ratio  $\nu$ , density  $\rho$ , film thickness  $d$  (Schneider and Schwarz 1997, reprinted with permission of Elsevier, original Fig. 1)

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \frac{E}{2(1+\nu)} \frac{\partial^2 u_i}{\partial x_i^2} + \frac{E}{2(1+\nu)(1-2\nu)} \frac{\partial^2 u_k}{\partial x_i \partial x_k} \quad (i, k = 1, 2, 3) \quad (3)$$

It can be shown that this system of three equations has two independent solutions of bulk waves in an infinitive medium, longitudinal and transverse waves.

The phase velocities  $c_l$  and  $c_t$  (Eqs. 4 and 5) of both wave modes depend on Young's modulus  $E$ , Poisson's ratio  $\nu$ , and density  $\rho$ .

$$c_l = \sqrt{\frac{E(1-\nu)}{2\rho(1+\nu)(1-2\nu)}} \quad (4)$$

$$c_t = \sqrt{\frac{E}{2\rho(1+\nu)}} \quad (5)$$

The question arises, if waves can be solutions of the equation of wave motion, which propagate along a free surface of a half-infinite medium and do not lose wave energy into the bulk. For the displacement  $u$  of such a surface wave, partial waves with longitudinal and transverse wave motion are combined in Eqs. 6, 7, and 8 (Farnell 1970).

$$u_1 = \{A^l \alpha_1^l \exp(ik_3^l x_3) + A^t \alpha_1^t \exp(ik_3^t x_3)\} \exp[ik(x_1 - ct)] \quad (6)$$

$$u_2 = 0 \quad (7)$$

$$u_3 = \{A^l \alpha_3^l \exp(ik_3^l x_3) + A^t \alpha_3^t \exp(ik_3^t x_3)\} \exp[ik(x_1 - ct)] \quad (8)$$

$u_1$  is the displacement of the particle along the surface in the direction  $x_1$  (Fig. 2),  $u_2$  is the displacement perpendicular to the sagittal plane, and  $u_3$  is the displacement into the material.  $k$  is the wave vector of the assumed surface acoustic wave,  $c$  is the phase velocity, and  $\omega$  is the circular frequency. The first and the second term in Eqs. 6 and 8 represent the longitudinal and the transverse contribution to the wave motion of the surface acoustic wave. The phase velocity  $c$  is defined to depend on  $k$  and  $\omega$ ,  $c = \omega/k$ . It is to note that  $k$ ,  $c$ , and  $\omega$  are invariant for both the longitudinal and transverse displacement in the surface wave motion in reference to the direction  $x_1$  along the surface.  $k_3^l$  and  $k_3^t$  are the components of the wave vectors of the longitudinal and the transverse displacement in the direction perpendicular to the surface into the material.  $\alpha_i^l$  and  $\alpha_i^t$  are the components of the corresponding eigenvectors along ( $i = 1$ ) and perpendicular ( $i = 3$ ) to the surface.  $A^l$  and  $A^t$  are amplitude coefficients which specify the relative contribution of the partial waves to the surface wave. They can be calculated from the boundary condition apart from a remaining arbitrary amplitude factor.

Eqs. 6 and 8 suggest that a surface acoustic wave exists when the wave equation has complex solutions for  $k_3^l$  and  $k_3^t$ . Only in this case, the exponential functions of the longitudinal and transverse amplitudes decrease in the  $x_3$  direction. Inserting the wave function (6) and (8) into the equation of wave motion (Eq. 3) provides a system of linear equations whose solvability depends on the zeros of the secular equation (characteristic polynomial). For isotropic materials and crystals with planes of high symmetry, the characteristic polynomial is bi-cubic and has four roots. The two roots, which meet the requirements that the amplitude decays in the  $x_3$ -direction, are given in Eqs. 9 and 10 (Farnell 1970).

$$k_3^l = i \cdot k \cdot b^l = i \cdot k \sqrt{1 - \left(\frac{c}{c_l}\right)^2} \quad (9)$$

$$k_3^t = i \cdot k \cdot b^t = i \cdot k \sqrt{1 - \left(\frac{c}{c_t}\right)^2} \quad (10)$$

$k_3^l$  and  $k_3^t$  are related to the wave vector  $\mathbf{k}$  of the surface wave with specific scaling factors  $b^l = \sqrt{1 - (c/c_l)^2}$  and  $b^t = \sqrt{1 - (c/c_t)^2}$ . The requirements  $b_l > 0$  and  $b_t > 0$  have the consequence that the phase velocity  $c$  of the surface acoustic wave has to be lower than the velocities of the bulk waves,  $c_l$  and  $c_t$ . Making use of the boundary conditions for the stress at the surface enables the phase velocity  $c$  of the surface acoustic wave to be determined. The following boundary conditions can be formulated, if the surface acoustic wave propagates in the  $x_1$ -direction.

$$\sigma_{31} = 0 \text{ at the surface} \quad x_3 = 0 \quad (11)$$

$$\sigma_{33} = 0 \text{ at the surface} \quad x_3 = 0 \quad (12)$$

These boundary conditions form a system of two equations, which has the following characteristic polynomial (Farnell 1970).

$$\left[2 - \left(\frac{c}{c_t}\right)^2\right]^2 - 4 \left[1 - \left(\frac{c}{c_l}\right)^2\right]^{1/2} \left[1 - \left(\frac{c}{c_t}\right)^2\right]^{1/2} = 0 \quad (13)$$

The suitable root for the phase velocity  $c$  of the surface acoustic wave has to be found numerically. A good analytical approximation is given in Eq. 1. It is also known in the form of Eq. 14 which illustrates the relation between the phase velocity  $c$  of the surface acoustic wave and the phase velocity  $c_t$  of the shear wave (Farnell 1970).

$$\frac{c}{c_t} = \frac{0,87 + 1,12\nu}{1 + \nu} \quad (14)$$

In the past, the interest in surface acoustic waves was mainly driven by very different fields, such as seismology, sensor applications, and the signal processing, e.g., in surface acoustic wave filters. Recently, the nondestructive testing community discovered the surface acoustic waves for the characterization of surfaces and coated materials. The fundamental work in seismology dates back to the 1950s of the last century. The shock waves of earthquakes were found to have high content of surface waves. Since the earth is a large body consisting of layers of very different properties, the theoretical works concentrated on the description of surface acoustic waves in layered materials. The review article of Lowe (1995) distinguishes two theoretical approaches for surface acoustic waves in layered materials. The transfer matrix method, introduced by Thomson (1950) and Haskell (1953), enables a relatively fast mathematical solution for multilayered surfaces, but it becomes unstable when the frequency, the thickness, or the number of layers become large. The alternative approach is the global matrix method (Knopoff 1964; Farnell and Adler 1972) whose solution procedure is more stable, but with increasing number of layers the size of the matrix becomes large and the solution becomes distinctly slower.

For a single layer on a substrate, the theory formulates four additional boundary conditions for the interface, given in Eqs. 15, 16, 17, and 18. The compression stress  $\sigma_{33}$  and the transverse shear stress  $\sigma_{31}$  are continuous at the interface between layer and substrate, and likewise the longitudinal particle displacement  $u_1$  and the vertical particle displacement  $u_3$  are continuous at the interface. The thickness of the layer is  $d$ , the zero point of  $x_3$  is at the surface of the layer (Farnell and Adler 1972).

$$\sigma_{33}^{\text{layer}} = \sigma_{33}^{\text{Sub}} \quad \text{at the interface } x_3 = d \quad (15)$$

$$\sigma_{31}^{\text{layer}} = \sigma_{31}^{\text{Sub}} \quad \text{at the interface } x_3 = d \quad (16)$$

$$u_1^{\text{layer}} = u_1^{\text{Sub}} \quad \text{at the interface } x_3 = d \quad (17)$$

$$u_3^{\text{layer}} = u_3^{\text{Sub}} \quad \text{at the interface } x_3 = d \quad (18)$$

Together with the requirement that the surface is stress-free, Eqs. 11 and 12, the boundary conditions form a system of equations of six rows. Wave functions similar to those in Eqs. 6 and 8 are introduced for the layer and the substrate (Farnell and Adler 1972). Whereas for the substrate only the roots  $k_{3,\text{Sub}}^l = i \cdot k \cdot b_{\text{Sub}}^l$  and  $k_{3,\text{Sub}}^t = i \cdot k \cdot b_{\text{Sub}}^t$  (Eqs. 9 and 10) which decrease the amplitude have to be taken into account, in the layer the amplitude can both decrease and increase. Consequently, the wave function for the layer is formed with four partial waves which have the  $k_3$ -vectors  $k_{3,\text{Layer}}^l = i \cdot k \cdot b_{\text{Layer}}^l$ ,  $k_{3,\text{Layer}}^t = -i \cdot k \cdot b_{\text{Layer}}^l$ ,  $k_{3,\text{Layer}}^l = i \cdot k \cdot b_{\text{Layer}}^t$  and  $k_{3,\text{Layer}}^t = -i \cdot k \cdot b_{\text{Layer}}^t$ . The solution has a total of six partial waves, two in the substrate,  $A_{\text{Sub}}^{l,\text{dec}}$  and  $A_{\text{Sub}}^{t,\text{dec}}$ , and four in the layer with the amplitudes  $A_{\text{Layer}}^{l,\text{dec}}$ ,  $A_{\text{Layer}}^{l,\text{inc}}$ ,

$A_{\text{Layer}}^{t,\text{dec}}$ ,  $A_{\text{Layer}}^{t,\text{inc}}$ . For the substrate, the displacements  $u_1^{\text{Sub}}$  and  $u_2^{\text{Sub}}$  are given in Eqs. 6 and 8. In the same way,  $u_1^{\text{layer}}$  and  $u_2^{\text{layer}}$  are formed from the sum of the four partial waves of the layer. Inserting  $u_1^{\text{Sub}}$ ,  $u_2^{\text{Sub}}$ ,  $u_1^{\text{layer}}$ , and  $u_2^{\text{layer}}$  into the boundary conditions of Eqs. 11, 12, and 15, 16, 17, and 18 yields a linear system of equations with six rows and six columns (Farnell and Adler 1972). The coefficients of this matrix are functions of the wave vector  $k$  and the phase velocity  $c$  of the surface acoustic wave. Furthermore, they depend on the material parameters Young's modulus  $E$ , Poisson's ratio  $\nu$ , and density  $\rho$  of both the layer and substrate which are introduced by the bulk wave velocities  $c_l$  and  $c_t$  of both materials and on the layer thickness  $d$ . If there is a solution for the surface acoustic wave, the determinant of the coefficient matrix must be zero. The exact solution has to be found by a numerical algorithm. Assuming the material parameters to be known, the calculation yields, for a given wave vector  $k$ , the related phase velocity  $c$  of the surface acoustic wave. This is the dispersion relation  $c = c(k)$ . The wave vector  $k$  and the film thickness  $d$  occur only in the product  $d \cdot k$ , hence the dispersion relation is sometimes written in the form  $c = c(d \cdot k)$  or  $c = c(d/\lambda)$  taking into account that  $k = 2\pi/\lambda$ . Many experimental setups measure the phase velocity  $c$  as it depends on the frequency  $f$ , which is termed dispersion curve  $c(f)$ . Therefore, the wave vector  $k$  is replaced by the frequency  $f$  to fit the theory to an experimental curve. The dispersion relation with all parameters involved can then be written in the general form (Eq. 19).

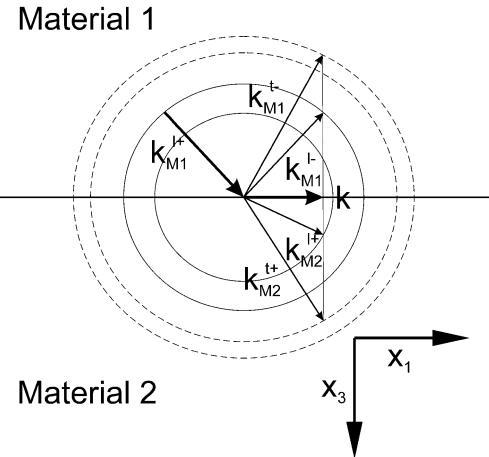
$$c = c(E_{\text{Sub}}, \nu_{\text{Sub}}, \rho_{\text{Sub}}, E_{\text{Layer}}, \nu_{\text{Layer}}, \rho_{\text{Layer}}, d, f) \quad (19)$$

The dimension of the matrix of the boundary conditions increases with increasing numbers of layers. Xiao and You (2006) presented the solution for two layers on a substrate, for which the matrix has the dimension of  $10 \times 10$ . In general, if  $n$  denotes the number of layers, the matrix of the boundary conditions has the dimension  $2(2n + 1)$ .

An alternative theoretical approach for solving the problem of propagation of surface acoustic waves in layered materials is the transfer matrix method, comprehensively described by Lowe (1995). It surveys incident and reflected longitudinal and shear waves at the boundary between two materials.

Figure 3 shows an incident longitudinal wave vector  $k_{M1}^{l+}$  splitting into four wave modes, the reflected longitudinal wave  $k_{M1}^{l-}$ , the reflected shear wave  $k_{M1}^{t-}$ , the transmitted longitudinal wave  $k_{M2}^{l+}$ , and the transmitted shear wave  $k_{M2}^{t+}$ . Reflection and transmission of waves at a boundary is described by Snell's law which requires that for the interaction of the partial waves, they must have the same frequency and the same wave vector component  $k$  in the direction  $x_1$  of the interface. If the wave propagates through a layer stack of successive materials 1 to  $n$ , this condition applies for all interfaces of the stack. Therefore, the wave vector component  $k$  along the interfaces is an invariant of the wave propagation in all layers. All field equations for all locations in all layers have the factor  $\exp(ikx_1)$ . In the case that the wave is a surface acoustic wave,  $k$  is the wave vector of this wave mode and this factor is  $\exp[ik(x_1 - ct)]$ , which also occurs in the partial wave approach of Farnell (1970) in Eqs. 6 and 8.

**Fig. 3** Wave vectors of reflected and transmitted acoustic waves at the interface between two materials



The model of the transfer matrix method (Thomson 1950; Haskell 1953; Lowe 1995) for the wave motion in a multilayer assumes also the superposition of longitudinal and shear bulk waves in each layer. The transfer of the wave motion from one layer to the other is determined by the boundary conditions at the interface for the displacement,  $u_1$  and  $u_3$ , and the stresses,  $\sigma_{33}$  and  $\sigma_{31}$ , already introduced in Eqs. 15, 16, 17, and 18. Four bulk waves with the following amplitudes are considered:  $A^{l+}$  and  $A^{t+}$  are the longitudinal and shear wave amplitudes incident to the interface, and  $A^{l-}$  and  $A^{t-}$  are the longitudinal and shear wave amplitudes which reflect from the interface. The displacement vectors  $u^l$  and  $u^t$  of the longitudinal and the shear wave have the form in Eqs. 21 and 22.  $u^l$  and  $u^t$  are the solutions of the wave equation with the restriction to propagate only along the direction  $x_1$  of interfaces of the layer-stack with the invariant wave vector  $k$  and the surface wave velocity  $c$ , since they decrease exponentially in the  $x_3$  direction. In the transfer matrix model, the scaling factors  $b^l$  and  $b^t$  in Eqs. 21 and 22 are derived from Snell's law, which yields the same relations as given in Eqs. 9 and 10.

$$u^l = \begin{pmatrix} k \\ 0 \\ b^l k \end{pmatrix} A^l \exp[ik(x_1 + ib^l x_3 - ct)] \quad (21)$$

$$u^t = \begin{pmatrix} k \\ 0 \\ -b^t k \end{pmatrix} A^t \exp[ik(x_1 + ib^t x_3 - ct)] \quad (22)$$

Making use of Eqs. 21 and 22 enables the displacements and the stresses at the interface to be calculated as they depend on the incident and reflected longitudinal and shear waves, given in Eq. 23.

$$\begin{pmatrix} u_1 \\ u_3 \\ \sigma_{33} \\ \sigma_{31} \end{pmatrix} = F \cdot \begin{pmatrix} A^{l+} \\ A^{l-} \\ A^{t+} \\ A^{t-} \end{pmatrix} = \begin{bmatrix} -k & -k & ib^t k & -ib^t k \\ -ib^l k & ib^l k & -k & -k \\ 2ib^l k & -2ib^l k & (1 + (b^t)^2)k & (1 + (b^t)^2)k \\ -(1 + (b^t)^2)k & -(1 + (b^t)^2)k & 2ib^t k & -2ib^t k \end{bmatrix} \cdot \begin{pmatrix} A^{l+} \\ A^{l-} \\ A^{t+} \\ A^{t-} \end{pmatrix} \quad (23)$$

The field matrix  $F$  with a  $4 \times 4$  dimension transfers the vector of the bulk wave amplitudes  $A^{l+}, A^{l-}, A^{t+}, A^{t-}$  into a vector of the displacements,  $u_1$  and  $u_3$ , and the stresses,  $\sigma_{33}$  and  $\sigma_{31}$  at the interface. In the case that the interface is a free surface along which a surface acoustic wave propagates, the surface is free of stress,  $\sigma_{33} = \sigma_{31} = 0$ , and the wave does not leak energy into the bulk material,  $A^{l-} = A^{t-} = 0$ . Inserting these requirements into Eq. 23 yields the following sub-system of equations:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{bmatrix} 2ib^l k & (1 + (b^t)^2)k \\ -(1 + (b^t)^2)k & 2ib^t k \end{bmatrix} \cdot \begin{pmatrix} A^{l+} \\ A^{t+} \end{pmatrix} \quad (24)$$

There is a non-trivial solution for  $A^{l+}$  and  $A^{t+}$ , if the determinant of the  $2 \times 2$  matrix has zeros. Introducing  $b^l = \sqrt{1 - (c/c_l)^2}$  and  $b^t = \sqrt{1 - (c/c_t)^2}$  from Eqs. 9 and 10 yields the same characteristic polynomial for the computation of the phase velocity  $c$  of the surface acoustic wave as derived in Eq. 13.

In the case that the interface connects two layers  $L1$  and  $L2$  of different materials located in the plane at  $x_3 = 0$ , the boundary conditions in Eqs. 15, 16, 17, and 18 apply which can be written in the form of Eq. 25.

$$\begin{pmatrix} u_1 \\ u_3 \\ \sigma_{33} \\ \sigma_{31} \end{pmatrix}_{L1, x_3=0} = \begin{pmatrix} u_1 \\ u_3 \\ \sigma_{33} \\ \sigma_{31} \end{pmatrix}_{L2, x_3=0} \quad (25)$$

Assuming the field at the interface  $x_3 = 0$  in layer  $L1$  is known and, therefore, also in  $L2$ , according to Eq. 25, the wave amplitudes  $A^{l+}, A^{l-}, A^{t+}$ , and  $A^{t-}$  can be calculated for layer  $L2$  with Eq. 27 by inverting Eq. 23 and introducing the phase matrix  $P$  in Eq. 26. In this way the wave field can be calculated at any depth of layer  $L2$ . In Eqs. 26 and 27, this depth is equal to the thickness of the layer  $d_2$ .

$$P = \begin{bmatrix} \exp(ib_{L2}^l kd_2) & 0 & 0 & 0 \\ 0 & \exp(-ib_{L2}^l kd_2) & 0 & 0 \\ 0 & 0 & \exp(ib_{L2}^t kd_2) & 0 \\ 0 & 0 & 0 & \exp(-ib_{L2}^t kd_2) \end{bmatrix} \quad (26)$$

$$\begin{pmatrix} A^{l+} \\ A^{l-} \\ A^{t+} \\ A^{t-} \end{pmatrix}_{L2, x_3=d_2} = P_{L2} \cdot F_{L2}^{-1} \cdot \begin{pmatrix} u_1 \\ u_3 \\ \sigma_{33} \\ \sigma_{31} \end{pmatrix}_{L2, x_3=0} \quad (27)$$

Making use of both Eqs. 23 and 27 allows the transfer of displacements and stresses from the top of the layer  $L2$  to the bottom of the layer  $L2$ , according to Eq. 28.

$$\begin{pmatrix} u_1 \\ u_3 \\ \sigma_{33} \\ \sigma_{31} \end{pmatrix}_{L2, x_3=d_2} = F_{L2} \cdot \begin{pmatrix} A^{l+} \\ A^{l-} \\ A^{t+} \\ A^{t-} \end{pmatrix}_{L2, x_3=d_2} = F_{L2} \cdot P_{L2} \cdot F_{L2}^{-1} \cdot \begin{pmatrix} u_1 \\ u_3 \\ \sigma_{33} \\ \sigma_{31} \end{pmatrix}_{L2, x_3=0} \quad (28)$$

Multiplying the field and phase matrixes in Eq. 28 yields the transfer matrix  $T_{L2}$  for the single layer  $L2$ .

$$T_{L2} = F_{L2} \cdot P_{L2} \cdot F_{L2}^{-1} \quad (29)$$

The boundary conditions require again that the displacements and the stresses at the bottom of the layer  $L2$  (Eq. 28) are equal to those at the top of the next layer  $L3$ . This process can be continued layer by layer for all subsequent layers, resulting in the transfer matrix  $S$  of the whole layer stack:

$$S = T_{L1} \cdot T_{L2} \cdots T_{Ln} \quad (30)$$

It is assumed that on the top of the first layer  $L1$  is a free surface, at  $x_3 = 0$ . Furthermore, it is assumed that on the bottom the layer stack is connected to an infinite substrate with the field matrix  $F_{\text{Sub}}$ . Finally, a surface acoustic wave propagating along the surface which does not leak energy into the substrate is assumed. These conditions are expressed in the following:

$$\begin{aligned} \begin{pmatrix} u_1 \\ u_3 \\ 0 \\ 0 \end{pmatrix}_{x_3=0} &= CS \cdot \begin{pmatrix} 0 \\ A^{l-} \\ 0 \\ A^{t-} \end{pmatrix}_{x_3=d_1+d_2+\dots+d_n} \\ &= S \cdot F_{\text{Sub}} \cdot \begin{pmatrix} 0 \\ A^{l-} \\ 0 \\ A^{t-} \end{pmatrix}_{x_3=d_1+d_2+\dots+d_n} \end{aligned} \quad (31)$$

The zeros on the left side of Eq. 31 express the fact that the surface is free of stress,  $\sigma_{33} = \sigma_{31} = 0$ . The zeros on the right side of Eq. 31 indicate that waves propagating down into the substrate do not exist,  $A^{l+} = A^{t+} = 0$ . Multiplying the

$4 \times 4$  matrices  $S$  and  $F_{\text{Sub}}$  to calculate the matrix  $CS$  and expanding the relation to the rows with zero stress yields the following  $2 \times 2$  homogenous system of equations (Eq. 32).

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{bmatrix} CS_{32} & CS_{34} \\ CS_{42} & CS_{44} \end{bmatrix} \cdot \begin{pmatrix} A^{l-} \\ A^{t-} \end{pmatrix} \quad (32)$$

There are solutions for  $A^{l-}$  and  $A^{t-}$  which are not zero, if it is possible to satisfy the following requirement:

$$CS_{32} \cdot CS_{44} - CS_{34} \cdot CS_{42} = 0 \quad (33)$$

The coefficients  $CS_{ij}$  are functions of the bulk wave velocities  $c_l^{Li}$ ,  $c_t^{Li}$  and thickness  $d_i$  of the layers  $Li$ , the bulk wave velocities  $c_l^{\text{Sub}}$ ,  $c_t^{\text{Sub}}$  of the substrate, the wave vector  $k$  and phase velocity  $c$  of the surface acoustic wave. The bulk velocities  $c_l^{Li}$  and  $c_t^{Li}$  depend on the elastic parameters and the density of the material of layer  $Li$ . The bulk velocities  $c_l^{\text{Sub}}$ ,  $c_t^{\text{Sub}}$  depend on the elastic parameters and the density of the substrate. The wave vector  $k$  can be replaced by the frequency  $f$ , because experimental devices usually measure the dispersion curve in the form  $c = c(f)$ . If the material parameters are known, the phase velocity  $c$  of the surface acoustic wave can be computed for a given frequency  $f$  by finding numerically the zeros of Eq. 33.

$$c = c(E_{\text{Sub}}, \nu_{\text{Sub}}, \rho_{\text{Sub}}, E_{L1}, \nu_{L1}, \rho_{L1}, d_1, \dots, E_{Ln}, \nu_{Ln}, \rho_{Ln}, d_n, f) \quad (34)$$

The surface acoustic wave is termed modal solution of the transfer matrix model (Lowe 1995). It represents that wave mode which can propagate without external forces in a multilayer connected to a substrate material.

---

## Laser Generation of Surface Acoustic Waves

The first generation of ultrasound by laser pulses was described in 1963 by White (1963). Five years later, Lee and White (1968) published the generation of surface acoustic waves by laser. Meanwhile, the laser became quite widespread as a tool for generating acoustic waves in research and industry. Laser generation of sound has some advantages compared to the traditional techniques (Lyamshev 1981). It can be used remotely and is flexible over long distances without direct contact to the medium in which the acoustic wave field has to be excited. The waves can be created in a wide range of frequencies and with adapted directivity of the wave field, which can be controlled by the laser pulse length and the shape of the laser focus. The laser-acoustic source can be positioned very precisely and very fast, even on complex geometries.

Several mechanisms of the interaction of the laser beam with the materials for sound generation are known (Akhmanov and Gusev 1992): the thermoelastic excitation, the electrostrictive mechanism, laser initiation of the piezo-effect, sound

excitation by the processes of melting, and ablation in the material target. For nondestructive testing, the thermoelastic regime is most attractive, since the surface is not damaged. The thermoelastic effect is the result of the absorption of the laser light in a surface layer with the thickness of the absorption depth. With a suitable pulse laser, a short and rapid heating of this layer can be provoked, which generates an acoustic impulse that is transmitted into the material from the absorption layer. The wave amplitude depends linearly on the laser pulse intensity (Lyamshev 1981). This behavior enables the use of methods of linear signal processing for the laser-acoustic signals.

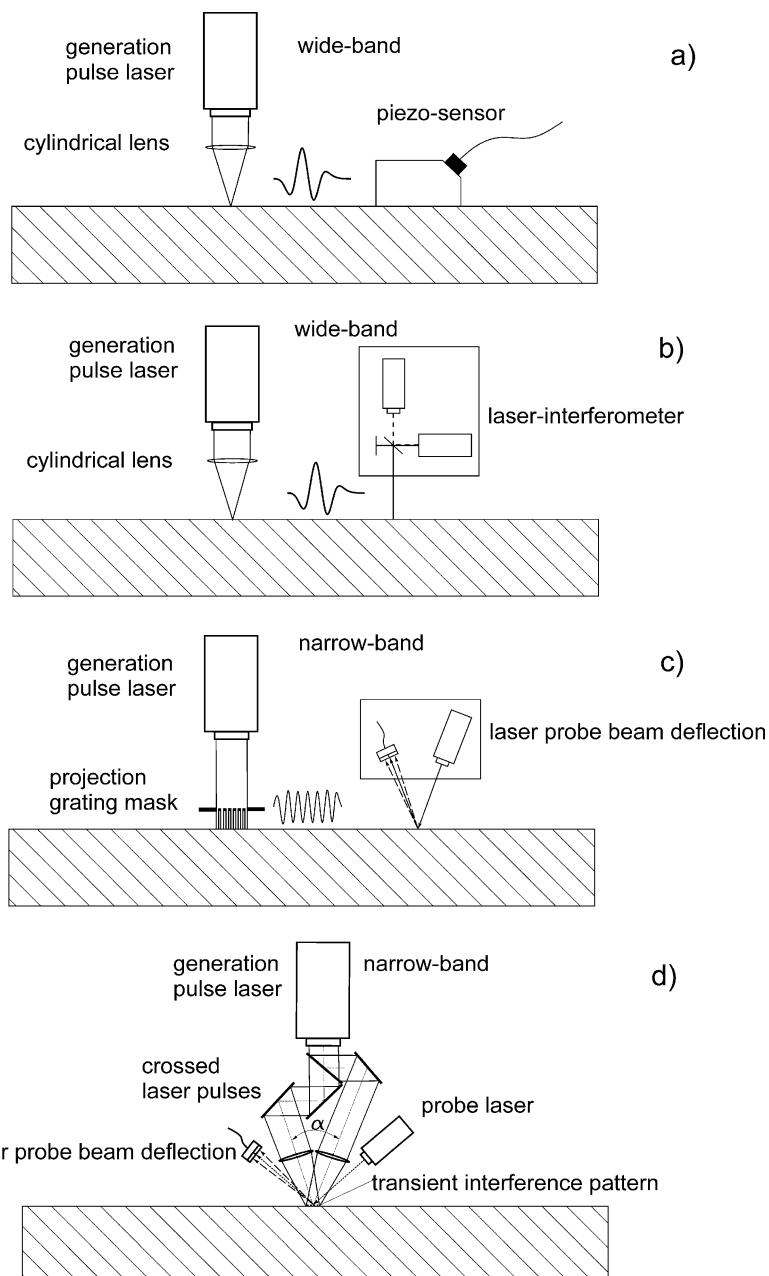
Lasers can generate all known elastic wave modes, e.g., longitudinal waves, shear waves, and surface acoustic waves. The effectiveness of the surface wave generation depends on the product of the wave vector  $k$  and the laser spot diameter  $D_L$ ,  $kD_L$ . It was found that for  $kD_L \approx 1$  the surface acoustic wave has a fraction of about 67% in all elastic waves generated, whereas the fraction of the shear wave is 26% and the fraction of the longitudinal wave only 7% (Karabutov 1985; Arnold et al. 1985). With increasing  $kD_L$  the fraction of the longitudinal wave intensity increases quickly and becomes dominant. The thermoelastic expansion of surface zone with larger diameter compared to the acoustic wavelength generates stress preferentially normal to the surface which radiates longitudinal waves into the bulk of the material. Therefore, generating surface acoustic waves of high frequencies requires a sharp focus of the laser beam on the additional condition that the laser beam is absorbed within a surface layer which is thin compared to the acoustic wavelength. If the surface acoustic wave is generated in a strongly light absorbing material, the shape of the acoustic impulse directly corresponds to the laser pulse (Karabutov 1985; Neubrand and Hess 1992). If the light penetration depth is large, the shape of the acoustic impulse depends only on the light absorption distribution over depth (Karabutov 1985). Focusing the laser beam by a cylindrical lens forms a laser-acoustic line source which transmits an approximately straight wave front within the near field of this line source. These wave fronts correspond to the “straight-crested” waves used in the theoretical approaches presented in section “[Surface Acoustic Waves](#),” which justifies analyzing experimental data by using this theory without diffraction correction. There are several publications which theoretically model a laser-acoustic line source which transmits surface acoustic waves (Karabutov 1985; Royer and Chenu 2000; Royer 2001). For example, Royer and Chenu (2000) calculated the acoustic wave field transmitted from a laser line with length  $L$ , using the time convolution of the laser pulse shape  $q(t)$  and a Green function  $G(r, t)$  in which the variable  $r$  is the distance between the observation point and a segment of line source. For the perpendicular distance  $x_1$  to the line where the wave field changes from the near to the far field, they estimated the relation in Eq. 35 ( $c$ : velocity of the surface acoustic wave,  $\tau$ : laser pulse duration).

$$x_1^N = \frac{L^2}{4c\tau} \quad (35)$$

Considering the experimental setup in Schneider and Schwarz (1997) with a length of the laser line of  $L = 4$  mm and laser pulse duration of  $\tau = 3$  ns, the far field begins at a distance  $x_1^N = 260$  mm from the laser line, if the sample is a (100) silicon wafer and measurement is done along the [110]-direction ( $c = 5080$  m/s). Whereas in the near field the wave front has a length that approximately corresponds to the length of the laser line, the wave front spreads sideway after entering into the far field and takes a circular shape. The surface acoustic wave amplitude decays with  $u \sim 1/\sqrt{R}$  ( $R$ : distance to the laser-acoustic line source), which is less than the bulk waves, because the bulk waves propagate as spherical waves in the far field (Karabutov 1985). It is recommended to measure the velocity of surface acoustic waves in the near field, because the diffraction loss is lower and the bending of the wave front does not influence the velocity measurement (Ruiz and Nagy 2002). It is useful also to take into account that the wave field for a given acoustic source depends on the frequency. Szabo and Slobodnik (1973) derived a formula for the extension of the near field depending on frequency  $f$ ,  $x_1^N = L^2 f / c$ . For an acoustic line source with a length of  $L = 4$  mm and surface wave velocity of  $c = 5080$  m/s ((100) silicon wafer, [110] direction), the near field length is  $x_1^N = 31.5$  mm for a frequency  $f = 10$  MHz and  $x_1^N = 315$  mm for  $f = 100$  MHz.

Figure 4a–d shows the schematic representations of four experimental setups which use laser-induced surface acoustic waves. The setups in Fig. 4a, b can measure the dispersion curve in a wide frequency range. The laser pulses are focused into a line onto the surface of the sample and generate acoustic impulses which are detected at a given distance by a piezoelectric sensor (Fig. 4a) or by a laser interferometer (Fig. 4b). Different types of sensors were used. Commercially available are bulk wave transducers, which are acoustically contacted to acryl glass wedges satisfying the critical angle for surface wave detection in the test material (Arnold et al. 1985; Schneider et al. 2012b). The wedge has to be coupled by a suitable medium to the sample surface. Coufal et al. (1992) have proposed a sensor that uses piezoelectric PVDF foils that can be contacted directly to the material surface without an acoustic coupling medium. This type of sensor can detect surface acoustic waves in a frequency range over 200 MHz. Interdigital transducers on piezoelectric substrate materials are commonly used in signal processing devices at very different frequencies (White 1970). It was demonstrated that these transducers can also detect surface acoustic waves in non-piezoelectric materials generated by laser pulses (Schneider and Franke 1990). The piezoelectric chip of the transducer must be coupled to the non-piezoelectric test sample by a drop of liquid. The angle between the transducer chip and the sample surface must be adjusted to a value that depends on the difference between the surface wave velocity of the chip and that of the sample material.

Combining the laser generation of surface acoustic waves with a laser interferometer offers the possibility of a completely non-contact measurement (Fig. 4b). Different types of laser interferometers were used for the detection of surfaces acoustic waves, based on the setup of Michelson (Neubrand and Hess 1992), Mach-Zehnder (Singer and Kufner 2017), Fabry-Perot (Shan et al. 1993).



**Fig. 4** Schematic representation of four laser-acoustic setups using surface acoustic waves for the characterization of material surfaces

An alternative technique to the relatively complicated interferometer setup is the laser-probe beam deflection method schematically shown in Fig. 4c (Lomonosov et al. 2001; Grünwald et al. 2015) whose principles are related to the knife-edge technique (Whitman and Korpel 1969). This method uses a continuous laser beam focused on the sample surface and detects the variation of the reflection angle when the surface acoustic wave passes the laser spot. The smaller the laser spot, the higher the frequency that can be detected. Normally, optical detection methods require a smooth surface with high reflectivity to achieve a sufficient signal-to-noise ratio of the acoustic signal. Impulses of surface acoustic waves with bandwidth up to 200 MHz could be received and analyzed (Kolomenskii et al. 1995). The Fabry-Perot interferometer also allows wave detection on rough surfaces (Monchaline 1985), even on surfaces with thermally sprayed coatings (Lima et al. 2005).

There are also laser-acoustic techniques that use narrow-band acoustic waves. Reducing the bandwidth enhances the signal-to-noise ratio of laser-acoustic signal. A grating of many periodic laser lines on the surface can emit a long wave track of narrow bandwidth. Such a grating can be realized by a projection mask with periodic slits that is positioned in the beam of the laser pulse near the surfaces of the sample (Fig. 4c) (Bennis et al. 2006). The distance between the periodic lines determines the wavelength  $\lambda$  and the number of the lines determines the bandwidth of the wave trace. The wave can be detected by the laser probe beam deflection technique. The spectral analysis of the signal yields the frequency  $f$  of the wave. The wave velocity  $c$  is calculated from  $c = \lambda f$  (Bennis et al. 2006). Changing the periodic line distance of the mask yields a measurement for another wavelength  $\lambda$  and, consequently, for another frequency  $f$ . In this way the dispersion curve  $c(f)$  can be measured in discrete steps.

A grating of laser lines can also be created by crossing the two beams split from a single laser pulse on the surface of the sample, as shown in Fig. 4d. In the plane of the surface, an interference line pattern is created with a periodic distance calculated in Eq. 36 which also determines the acoustic wavelength  $\lambda$  (Rogers et al. 1994).

$$\lambda = \frac{\Lambda}{2 \sin(\theta/2)} \quad (36)$$

In Eq. 36, the term  $\Lambda$  denotes the wavelength of the laser pulse and  $\theta$  is the angle between the two beams split from a single beam. Varying the cross angle  $\theta$  offers a flexible possibility of changing the period of the grating and, accordingly, the wavelength  $\lambda$ . Laser grating techniques are able to measure at frequencies higher than 1 GHz (Maznev et al. 2003).

---

## Experimental Setup and Measuring Procedure

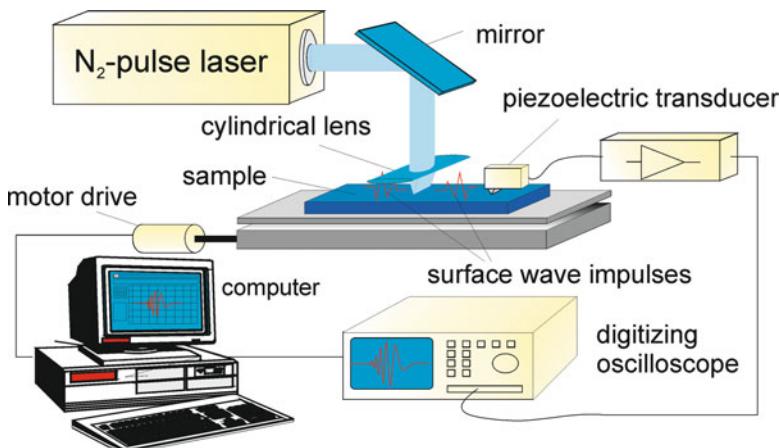
Several experimental setups were developed to measure the dispersion of surface acoustic waves which occur, if the material surfaces are coated with a film or are modified by surface treatments (Neubrand and Hess 1992; Schneider et al. 1992;

Lima et al. 2005; Xiao et al. 2016; Singer and Kufner 2017). In this chapter, an example of an experimental setup is described, which corresponds to the schematic representation in Fig. 4a. It uses a pulse laser to generate a wideband surface acoustic wave pulse and a piezoelectric sensor to detect the acoustic pulse form.

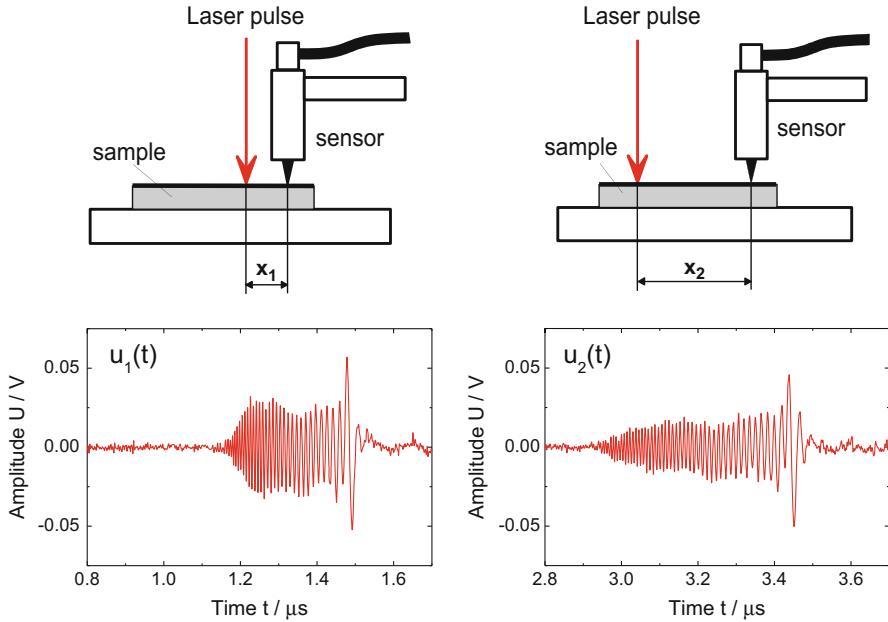
The schematic representation of the test equipment is shown in Fig. 5. A nitrogen pulse laser is used, with a laser wavelength of 337 nm, pulse energy of 90  $\mu$ J, pulse peak power of 30 kW, beam divergence  $\leq 0.5 \text{ } \mu\text{rad}$ , and pulse duration of 3 ns. The laser beam is focused by a cylindrical lens on the sample surface into a line with a length of 4 mm. The laser pulse generates a wideband surface wave impulse which is detected by a wideband piezoelectric transducer (bandwidth: 250 MHz) designed following the suggestions of Coufal et al. 1992. For rough surfaces and materials with high acoustic attenuation, a wedge-transducer is used, which is commercially available and has a frequency ranges from  $f = 1$  to 20 MHz.

Specimen and transducer are fixed on a translation stage (positioning error:  $\Delta x \leq 1.5 \text{ } \mu\text{m}$ ) that moves perpendicular to the position of the laser beam to vary precisely the distance  $x$  between laser focus line and transducer. The signals are recorded by an oscilloscope with a sampling rate higher than 2 GSa/s and are transferred to a computer. The computer controls the measuring procedure and performs the signal processing to calculate the dispersion curve  $c(f)$ . Surface acoustic wave impulses are generated at least at two different distances  $x_1$  and  $x_2$  between the laser focus line and the transducer.

An example of two laser-acoustic signals  $u_1(t)$  and  $u_2(t)$  is shown in Fig. 6. The signals  $u_1(t)$  and  $u_2(t)$  would have the same pulse form, if they were detected on the surface of a homogeneous material, because all waves independent on their frequency would propagate with the same velocity. The laser-acoustic signals in Fig. 6 were detected on a (100) silicon wafer which was coated with an amorphous



**Fig. 5** Schematic representation of the laser-acoustic device for material testing with surface acoustic waves (Schneider and Schwarz 1997, reprinted with permission of Elsevier, original Fig. 2)



**Fig. 6** Laser-acoustic signals detected at two different distances  $x_1$  and  $x_2$  between the laser focus line and the piezoelectric transducer in a silicon single crystal coated with a 3.32  $\mu\text{m}$  thick ta-C film

diamond-like carbon film (ta-C) with a thickness of 3.32  $\mu\text{m}$ . This film has a distinctly higher Young's modulus than the silicon substrate and consequently a higher wave velocity. The effect of the film causes the wave packet to disperse with increasing distance between the laser focus line and the acoustic sensor. The wave packet can be regarded as additive superposition of elastic waves with different frequencies. The waves with higher frequencies are more influenced by the film due to their lower penetration depth. They are received before the lower-frequency waves. The impulse deformation contains interesting information about the film material (Young's modulus, density, thickness). The impulses  $u_j(t)$  ( $j = 1$  and 2) are transformed from the time to the frequency domain by Fourier transformation which provides the complex spectral function  $U_j(f)$ .

$$U_j(f) = \int_{-\infty}^{\infty} u_j(t) \cdot \exp(i2\pi ft) \cdot dt \quad (37)$$

$U_j(f)$  contains amplitude and phase information. The magnitude of the complex vector  $U_j(f)$  provides the amplitude spectrum of the signal according to Eq. 38.

$$|U_j(f)| = \sqrt{\{\text{Re}[U_j(f)]\}^2 + \{\text{Im}[U_j(f)]\}^2} \quad (38)$$

$\text{Re}[U_i(f)]$  and  $\text{Im}[U_i(f)]$  are the real and the imaginary parts of  $U_j(f)$ .

The angle  $\Phi_j(f)$  of the complex spectral function  $U_j(f)$  in the complex plane can be calculated as

$$\Phi_j(f) = \arctan \left\{ \frac{\text{Im}[U_j(f)]}{\text{Re}[U_j(f)]} \right\} + n_j 2\pi \quad (39)$$

$\Phi_j(f)$  represents the phase spectrum of the laser-acoustic signal  $u_j(t)$ . The phase values  $\Phi_j(f)$  calculated by Eq. 39 have an ambiguity of  $n_j 2\pi$ , since the arctan function provides only values within the range  $-\pi/2 \leq \Phi \leq \pi/2$ . The procedure of deriving the correct phase values is described in Schneider et al. (1992). It requires the laser-acoustic signals to be detected at some additional positions  $x_j$  between the positions  $x_1$  and  $x_2$ . Their number depends on the difference between the phase and group velocity and is usually between three and five.

The phase spectra  $\Phi_1(f)$  and  $\Phi_2(f)$  enable the phase velocity  $c$  as a function of frequency  $f$  to be calculated for the distance between  $x_1$  and  $x_2$  (Eq. 40) (Sachse and Pao 1978)

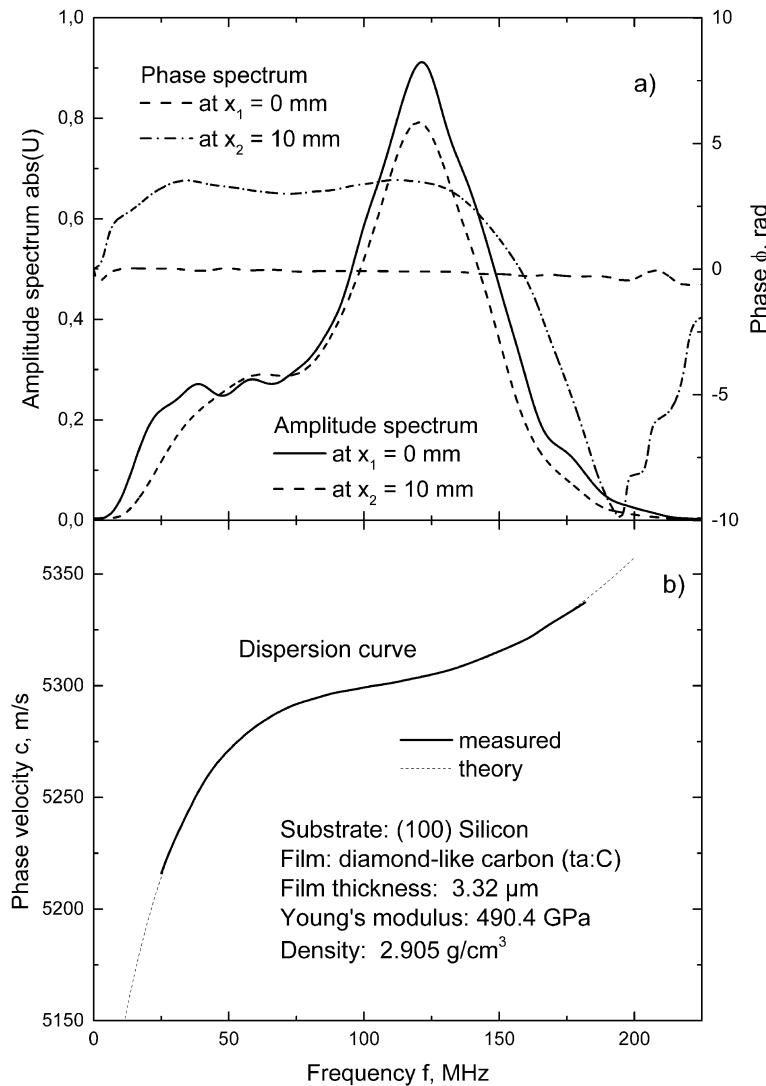
$$c(f) = \frac{(x_2 - x_1) 2\pi f}{\Phi_2(f) - \Phi_1(f)} \quad (40)$$

$c(f)$  is the dispersion curve of the surface acoustic wave.

Eq. 40 represents a difference measuring method. The signal  $u_1(t)$  at the position  $x_1$  yields the reference phase spectrum  $\Phi_1(f)$  for the instrument which is subtracted from the spectrum  $\Phi_2(f)$  of the signal  $u_2(t)$  at position  $x_2$ . In this way, only the phase shift on the measuring distance  $x_2 - x_1$  remains in  $\Phi_2(f) - \Phi_1(f)$  and the contributions of the test components, such as the laser, sensor, sensor coupling to the surface, signal amplifier, etc., are eliminated.

Figure 7a shows the amplitude and phase spectra of the two laser-acoustic signals  $u_1(t)$  and  $u_2(t)$  shown in Fig. 6. The amplitude spectrum at  $x_2 = 10$  mm is slightly lower than at  $x_1 = 0$  mm, revealing the attenuation of the wave over the measuring distance. The phase spectrum at  $x_1 = 0$  mm varies only little with the frequency  $f$ , because the zero point of the Fourier transformation was set to the signal maximum. The time difference between both signal maxima is converted into phase values and added to the difference of the phase spectra  $\Phi_2(f) - \Phi_1(f)$ .

Fig. 7b shows the dispersion curve calculated with Eq. 40. The theoretical curve fitted to the dispersion curve measured is also shown in the diagram. There is good agreement between the experimental and theoretical curves. The values of film thickness  $d = 3.32$  mm, Young's modulus  $E = 490.4$  GPa, and density  $\rho = 2.905$  g/cm<sup>3</sup> of the film, which were calculated by the fitting procedure are also given. The complex shape of the curve enables three film parameters to be calculated. Only the Poisson's ratio of the film  $\nu = 0.19$  was entered into the calculation, and the material parameters of the silicon substrate as the elastic constants  $C_{11} = 164.3$  GPa,  $C_{12} = 63.1$  GPa,  $C_{44} = 78.75$  GPa, and the density



**Fig. 7** (a) Amplitude and phase spectra of the two laser-acoustic signals shown in Fig. 6.  
(b) Dispersion curve calculated from the phase spectra in (a)

$\rho = 2.33 \text{ g/cm}^3$ . The experimental dispersion curve was only calculated for the frequency range where the amplitude is higher than 10% of the maximum of the spectrum. The theoretical curve can also be calculated beyond the bandwidth of the measurement, since the material parameters are known from the curve fit.

Ignoring the random error of the signal records, the uncertainty of this measurement can be estimated from the accuracy of time (oscilloscope) and distance (translation stage) measurement:  $\Delta t \leq \pm 0.1$  ns and  $\Delta x \leq \pm 1.5$   $\mu\text{m}$ . The noise effect was reduced by cross-correlating the signals. This procedure is described in more detail in Schneider and Schwarz (1997). Taking into account the signal-to-noise ratio (SNR) estimated from the amplitude spectrum with and without the signal, the bandwidth, and the correlation length enables the random error to be estimated (Aussel and Monchalin 1989). In the peak range of the amplitude spectrum, it does not exceed the accuracy of the time measurement. It is proportional to  $1/(\text{SNR}^{1/2}f)$  and, therefore, increases with decreasing frequency  $f$  and decreasing SNR. The highest random error of  $\Delta t \leq \pm 3$  ns was estimated at the lower limit ( $f = 25$  MHz) of the measured dispersion curve in Fig. 7b. For a measuring distance of  $x_2 - x_1 = 10$  mm, the velocity error can be estimated to  $\Delta c/c \leq \pm 2 \times 10^{-4}$  for the maximum range and to  $\Delta c/c \leq \pm 2.5 \times 10^{-3}$  for the range of the lower band limit.

The dispersion curve of the surface acoustic wave is used to calculate the material parameters of the surface film. The theoretical approaches described in section “[Surface Acoustic Waves](#)” have shown that the dispersion relation (Eq. 34) depends on Young’s modulus  $E$ , Poisson’s ratio  $\nu$ , density  $\rho$ , and film thickness  $d$ . The calculation of these material parameters requires the inverse solution of the dispersion relation. Because the dispersion relation is an implicit function of the material parameters, the method of nonlinear least-squares optimization (curve fitting) has to be used, formulated in Eq. 41.

$$\sum_l^N [c(E_{\text{Sub}}, \nu_{\text{Sub}}, \rho_{\text{Sub}}, E_{L1}, \nu_{L1}, \rho_{L1}, d_1, \dots, E_{Ln}, \nu_{Ln}, \rho_{Ln}, d_n, f_l) - c(f_l)]^2 \rightarrow \min \quad (41)$$

The first term is the theoretical curve computed from the dispersion relation and the second term represents the experimental data obtained from the measurement of  $c(f_l)$  with  $f_l$  being the discrete frequency values in the measured curve.  $N$  represents the total number of the data points which is typically in the range between one and two thousands. The Levenberg-Marquardt procedure is used for the least-squares optimization (Marquardt 1963). Usually, not all material parameters of the film and substrate involved in the theory can be fitted. The number of the parameters which can be calculated depends on the shape of the dispersion curve. A measured dispersion curve being a straight line can only yield two parameters, the Young’s modulus of the film from the slope of the curve and the Young’s modulus of the substrate from extrapolating the theoretical curve to the zero frequency. A dispersion curve with higher-order curvature can yield more film parameters, in the best case up to three parameters, Young’s modulus  $E$ , density  $\rho$ , and film thickness  $d$ , as shown in Fig. 7b. Both theories were applied, the transfer matrix method (Thomson 1950; Haskell 1953), the global matrix method (Farnell and Adler 1972), and yielded the same results.

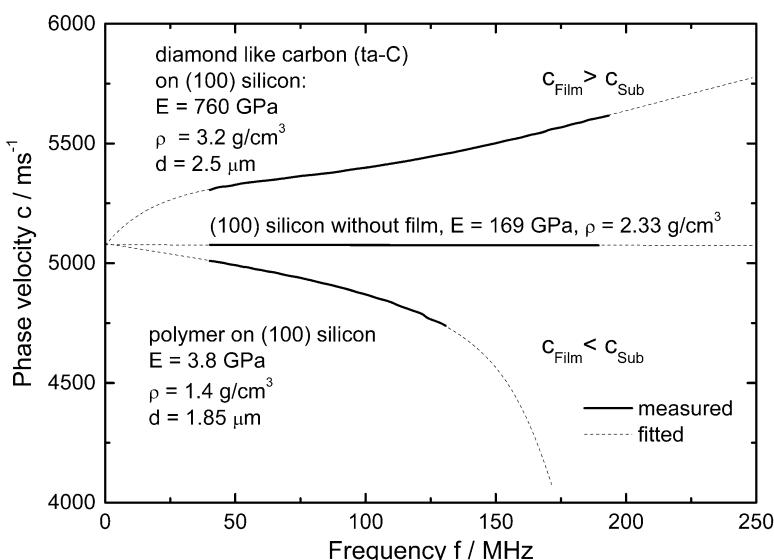
## Effect of Film and Substrate on the Surface Wave Dispersion

### The Effect of the Film Material

Figure 8 shows three important cases of surface wave dispersion (Schneider and Schwarz 1997). Two films with different properties were deposited on (100) silicon substrates, a super-hard diamond-like carbon film (ta-C) and a soft polymer film. Their effect on the surface acoustic wave dispersion is compared with the non-coated substrate. The velocity of the non-coated substrate is constant and does not show dispersion.

The ta-C-film with a thickness of 2.3  $\mu\text{m}$  has a considerably higher Young's modulus ( $E = 760 \text{ GPa}$ ) than the substrate ( $E = 169 \text{ GPa}$ ), whereas the density of the ta-C,  $\rho = 3.2 \text{ g/cm}^3$ , is not much higher than for silicon,  $\rho = 2.33 \text{ g/cm}^3$ . Consequently, the surface wave velocity of the film is higher than the velocity of the substrate,  $c_{\text{Film}} > c_{\text{Sub}}$ , which causes the phase velocity  $c$  to increase with frequency  $f$ . Increasing the frequency  $f$  reduces the penetration depth of the wave. The effect of the film on the wave propagation increases, whereas the effect of the substrate decreases. On the other hand, the polymer film being 1.85  $\mu\text{m}$  thick has a much lower Young's modulus of only  $E = 3.8 \text{ GPa}$ . This film has a distinctly lower surface acoustic wave velocity,  $c_{\text{Film}} < c_{\text{Sub}}$ . The velocity  $c$  decreases with frequency  $f$ .

Extrapolating the theoretical fit curves to the zero frequency shows that they meet at the same velocity at  $f = 0$ , the velocity of the substrate  $c_{\text{Sub}} = 5076 \text{ m/s}$ . For single



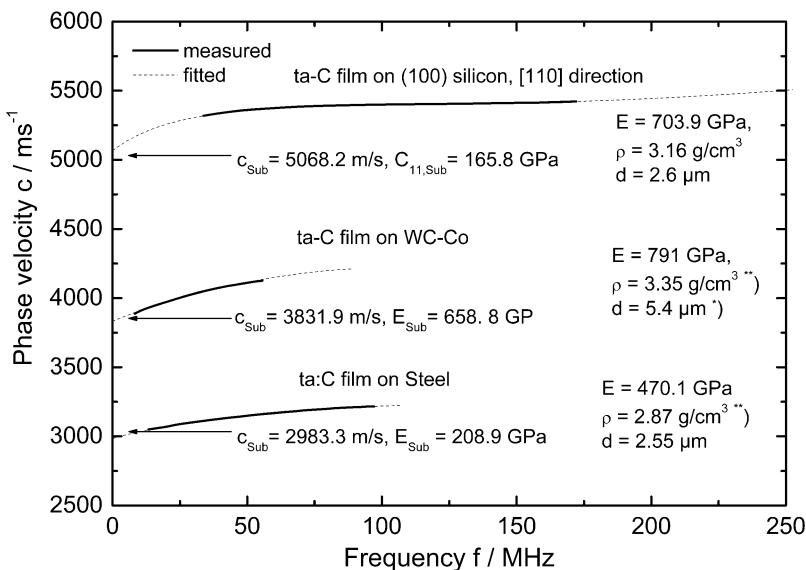
**Fig. 8** Three dispersion curves measured for (100) silicon coated with diamond-like carbon film (ta-C), coated with polymer film, and without coating; measuring direction: [110] (Schneider and Schwarz 1997, reprinted with permission of Elsevier, original Fig. 6)

crystals, the velocity  $c_{\text{Sub}}$  depends on the direction of the measurement. A measurement into the [100] direction on (100) silicon would result in a value of  $c_{\text{Sub}} = 4918 \text{ m/s}$ .

## The Effect of the Substrate Material

Figure 9 shows the effect of the substrate material on the dispersion curve (Schneider et al. 1997). Three different substrates, (100) silicon single crystal, cemented carbide WC-Co and steel, were coated with diamond-like carbon films (ta-C). The films were 2.6  $\mu\text{m}$ , 5.4  $\mu\text{m}$ , and 2.55  $\mu\text{m}$  thick, with the Young's moduli of  $E = 703.9 \text{ GPa}$ ,  $E = 791 \text{ GPa}$ , and  $E = 470.1 \text{ GPa}$ , respectively. Whereas the film determines the slope and the curvature of the dispersion curve, the substrate material determines the velocity  $c_{\text{Sub}}$  at  $f = 0$ . Changing the substrate material shifts the dispersion curve along the velocity axis.

Extrapolating the fit curves to the zero point yields the substrate velocity: for silicon  $c_{\text{Sub}} = 5068.2 \text{ m/s}$ , for WC-Co  $c_{\text{Sub}} = 3831.9 \text{ m/s}$ , for the steel  $c_{\text{Sub}} = 2983.3 \text{ m/s}$ . With the known density values of these materials (silicon  $\rho = 2.33 \text{ g/cm}^3$ , WC-Co  $\rho = 15.6 \text{ g/cm}^3$ , steel  $\rho = 7.8 \text{ g/cm}^3$ ), one elastic parameter of the substrate can be calculated by the fitting procedure in addition to the film parameters. For cubic silicon substrate, the elastic constant  $C_{11} = 165.8 \text{ GPa}$  was



**Fig. 9** Dispersion curves measured on (100) silicon, steel, and cemented carbide WC-Co all coated with diamond-like carbon films (ta-C), fitted results for the film parameters Young's modulus  $E$ , density  $\rho$ , and thickness  $d$  entered into the fitting; (\*\*\*) density  $\rho$  obtained from an empirical correlation between  $E$  and  $\rho$  (Schneider and Schwarz 1997, reprinted with permission of Elsevier, original Fig. 12)

calculated. For isotropic WC-Co and steel, the Young's moduli  $E = 658.8$  GPa and  $E = 208.9$  GPa were calculated. The remaining elastic parameters must be known and must be entered into the fit, for the cubic silicon, the elastic constants  $C_{12} = 63.5$  GPa and  $C_{44} = 79.6$  GPa, and for isotropic WC-Co and steel, the Poisson's ratio  $\nu = 0.19$  and  $\nu = 0.29$ , respectively.

The dispersion curves measured for the WC-Co and steel substrate have lower frequency limits, about 50 MHz and 100 MHz, compared to silicon that has an upper band limit of 220 MHz. This is due to the polycrystalline microstructure of the cemented carbide WC-Co and steel which causes scattering of the high-frequency waves at the grain boundaries. The lower bandwidth reduces the information content of the dispersion curve. The curve measured for the silicon substrate enabled independent calculation of three film parameters,  $E = 703.9$  GPa,  $\rho = 3.16$  g/cm<sup>3</sup>,  $d = 2.60$  μm. For the ta-C film on the steel substrate, Young's modulus  $E = 470$  and film thickness  $d = 2.55$  μm could be fitted. The density value  $\rho = 2.87$  g/cm<sup>3</sup> was obtained from a correlation between Young's modulus  $E$  and density  $\rho$  empirically found for hydrogen-free amorphous diamond-like carbon (Schneider et al. 1997). For the ta-C film on the cemented carbide WC-Co, only Young's modulus  $E = 791$  GPa could be fitted. The density  $\rho = 3.35$  g/cm<sup>3</sup> was also obtained from the empirical correlation between  $E$  and  $\rho$ , and the film thickness  $d = 5.4$  μm was measured with a stylus instrument.

## The Effect of Film Thickness

Figure 10 shows dispersion curves measured for diamond films deposited on (100) silicon (Schneider et al. 1997). The films had a thickness  $d$  from 0.08 to 5.5 μm. The curves ascend due to the higher velocity of the film compared to the substrate. The ascent is the steeper the thicker the film.

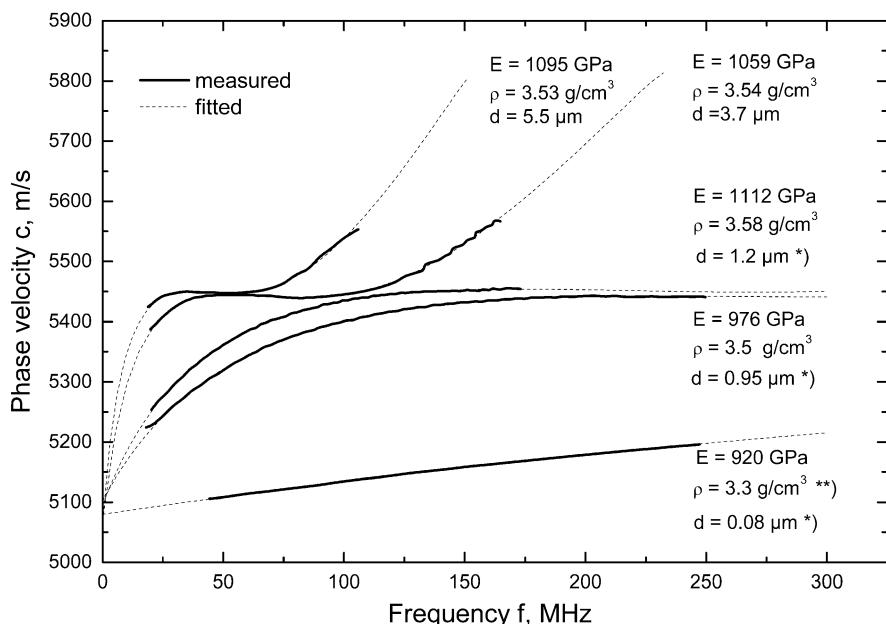
For the film thickness of  $d = 3.7$  μm and  $d = 5.5$  μm, the curves have a twofold curved shape with a point of inflection. Fitting the curves with the theory yields three material parameters, Young's modulus  $E$ , the density  $\rho$ , and the thickness  $d$  of the film. For a film thickness  $d = 0.95$  μm and  $d = 1.2$  μm, the point of inflection and the second ascent are beyond the upper limit of bandwidth of the measurement. Fitting these curves allows only two film parameters, Young's modulus  $E$  and the density  $\rho$ , to be calculated. The film thickness  $d$  was measured with a stylus instrument. For the thinnest film, the dispersion curve measured is a straight line. The regions which have significant curvatures for a multi-parameter fit are at much higher frequencies compared to the measuring range of the experimental setup. In this case only one parameter of the film can be calculated by the fit, Young's modulus  $E = 920$  GPa. Film thickness  $d$  and density  $\rho$  have to be entered into the computation. The film thickness was measured ellipsometrically. The density of this film was estimated from the correlation between Young's modulus  $E$  and the density  $\rho$  (Schneider et al. 1997). All values of Young's modulus  $E$  and the density  $\rho$  obtained for the films in Fig. 10 are within the range of the values known for polycrystalline diamond,  $E = 1143$  GPa and  $\rho = 3.52$  g/cm<sup>3</sup>. The curves of the thicker films have a lower bandwidth, due to the higher roughness of the surface.

The effect of a film on the surface wave dispersion depends on the ratio of the film thickness to the wavelength  $d/\lambda$ . For the surface acoustic wave, the wavelength  $\lambda$

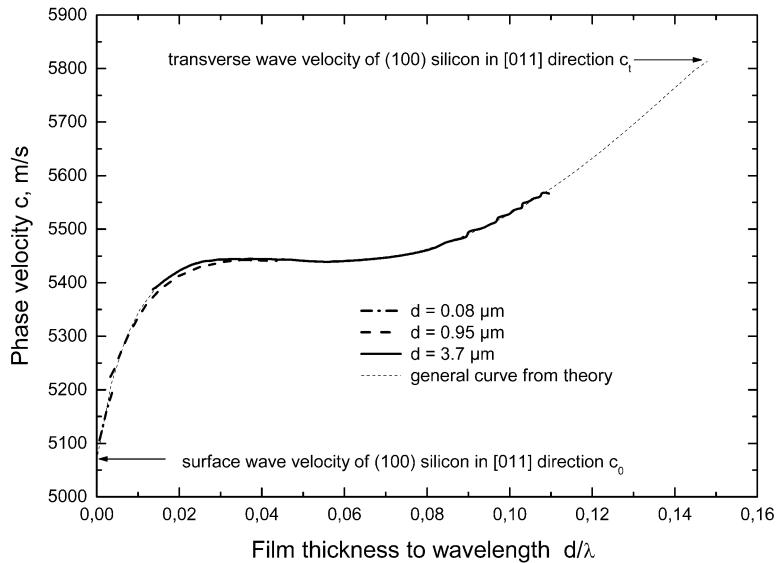
defines the penetration depth of the wave. The dependence of the dispersion curve on the frequency  $f$  can be replaced by the ratio of  $d/\lambda$ , dividing each frequency value  $f$  by the related phase velocity  $c$  and multiplying by the film thickness  $d$ ,  $d/c/f$ . In the normalized diagram shown in Fig. 11, the dispersion curves measured for different film thicknesses have to coincide with a general curve on the condition that all films have the same properties apart from differing in the film thickness.

The general curve in Fig. 11 shows the complete range of dispersion for this film-substrate combination. There is a cutoff ratio for  $d/\lambda$  at the velocity  $c_t$  of the transverse wave of the silicon substrate which is typical for ascending dispersion curves. In the  $c(f)$  curve, a cutoff frequency is observed. At the cutoff frequency the wave vector  $k_3^t$  in Eqs. 6 and 8 becomes real and the wave motion leaks into the bulk material as shear wave. Whereas the dispersion curves of thick films measured with the experimental setup in Fig. 5 end at the cutoff frequencies, other measuring techniques, such as Brillouin scattering, can also measure beyond the cutoff (Zinin et al. 2002).

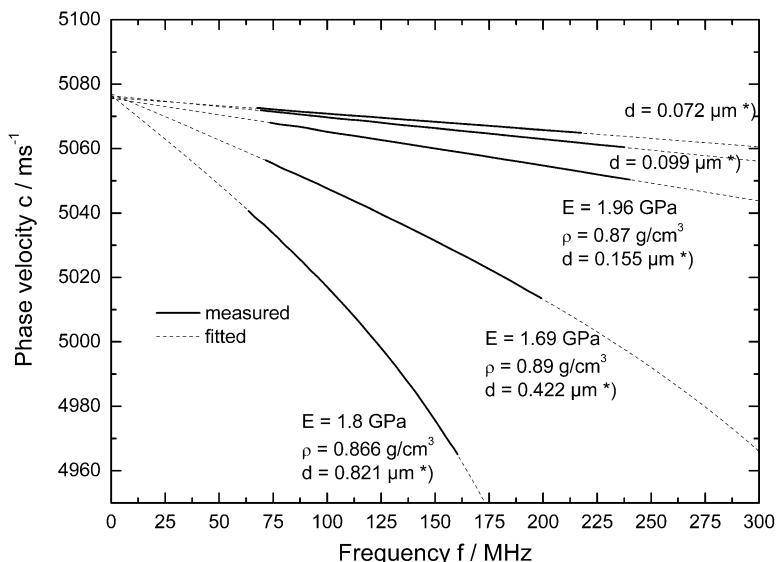
Figure 12 shows the effect of the film thickness for a film material which causes a descending dispersion curves. Silica xerogel SiCOH-films developed for low- $k$  applications in the semiconductor industry were deposited on (100) silicon (Schneider et al. 2005). These films with high porosity (up to 50%) had a very low Young's modulus, between  $E = 1.69$  and  $1.96$  GPa. The density of the films was between  $\rho = 0.866$  and  $0.89$  g/cm $^3$ . With increasing film thickness the descent of the



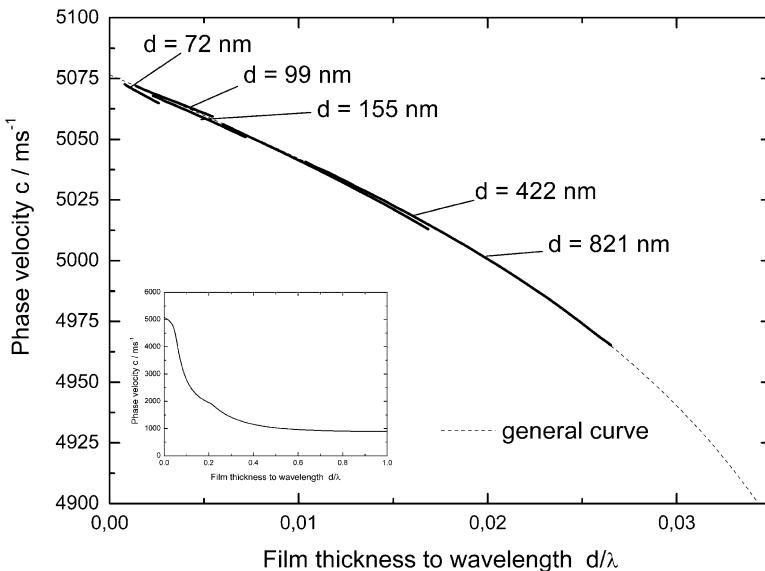
**Fig. 10** Dispersion curves measured on (100) silicon coated with diamond films, fit results for the film parameters Young's modulus  $E$ , density  $\rho$ , and thickness  $d$ ; \*) thickness  $d$  entered into the fit; \*\*) density  $\rho$  obtained from an empirical correlation between  $E$  and  $\rho$  (Schneider et al. 1997, reprinted with permission of Elsevier, original Fig. 5)



**Fig. 11** Normalized dispersion curves for diamond films on (100) silicon from Fig. 10: phase velocity  $c$  versus the ratio of film thickness to wavelength  $d/\lambda$  (Schneider et al. 1997, reprinted with permission of Elsevier, original Fig. 6)



**Fig. 12** Dispersion curves measured on (100) silicon coated with low- $k$  films, measuring direction [110], fit results for the film parameters Young's modulus  $E$ , density  $\rho$ , and thickness  $d$ ; \*) thickness  $d$  entered into the fit (Schneider et al. 2005, reprinted with permission of Elsevier, original Fig. 2)



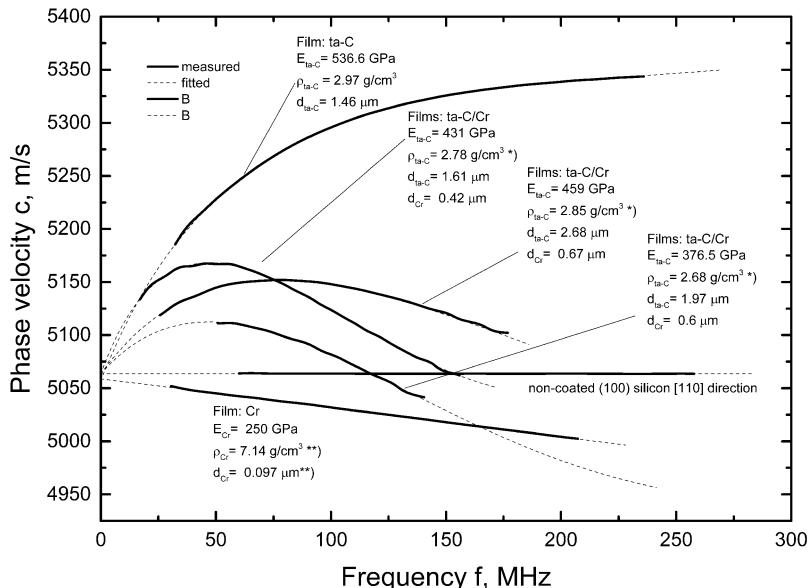
**Fig. 13** Normalized dispersion curves for low- $k$  films on (100) silicon from Fig. 12: phase velocity  $c$  versus the ratio of film thickness to wavelength  $d/\lambda$ , the complete dispersion curve is given in the inset diagram (Schneider et al. 2005, reprinted with permission of Elsevier, original Fig. 3)

dispersion curve becomes steeper. For the film thickness  $d > 155$  nm, the fit procedure allows two film parameters, Young's modulus  $E$  and the density  $\rho$ , to be calculated. The film thickness was measured ellipsometrically. The bandwidth of the dispersion curve reduces with increasing film thickness. This behavior is similar to the effect of diamond films. However, in this case the reason is the damping effect in the polymeric film material caused by the internal friction. The normalized diagram of the phase velocity  $c$  versus the ratio of film thickness to wavelength,  $d/\lambda$ , is shown for this film-substrate combination in Fig. 13.

The measured curves only slightly differ from the general theoretically calculated curve which indicates that the films have similar properties. The dispersion curves for a low velocity film on a higher velocity substrate do not have a cutoff, but approach asymptotically to the velocity of the film with increasing  $d/\lambda$ . These high porous films have a surface wave velocity of only 880 m/s. The complete curve is shown in the inset picture.

### The Effect of a Double Layer of Two Films with Different Properties

Figure 14 demonstrates the effect of a double layer combined of two very different films, a metal film of chromium and a super-hard ta-C film, on the dispersion of the surface acoustic wave.



**Fig. 14** Dispersion curves of ta-C films with a Cr interlayer on (100) silicon, ta-C/Cr, compared to the dispersion curves of the single layers of ta-C and Cr, measuring direction [110]; \*) density  $\rho_{ta-C}$  obtained from an empirical correlation between  $E$  and  $\rho$ ; \*\*) thickness  $d_{Cr}$  and density  $\rho_{Cr}$  entered into the fitting

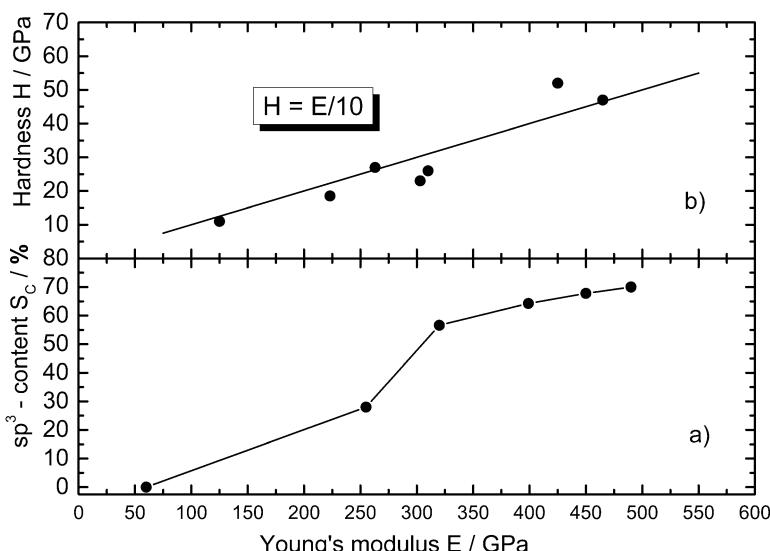
At first, single layers of ta-C with a thickness of  $d_{ta-C} = 1.46 \mu\text{m}$  and of chromium with a thickness of  $d_{Cr} = 0.097 \mu\text{m}$  were separately deposited on two (100) silicon substrates. The dispersion curve of the ta-C film increases with frequency. Three film parameters, Young's modulus  $E_{ta-C} = 536.6 \text{ GPa}$ , the density  $\rho_{ta-C} = 2.97 \text{ g/cm}^3$ , and the film thickness  $d_{ta-C} = 1.46 \mu\text{m}$ , could be calculated by the fitting procedure. The dispersion curve of the Cr-film decreases due to the higher density of the film  $\rho_{Cr} = 7.14 \text{ g/cm}^3$  compared to the silicon substrate,  $\rho_{Si} = 2.33 \text{ g/cm}^3$ . The curve is a straight line; therefore, only one film parameter could be fitted, Young's modulus of  $E_{Cr} = 250 \text{ GPa}$ . The film thickness  $d = 0.097 \mu\text{m}$  ellipsometrically measured and the tabulated density  $\rho_{Cr} = 7.14 \text{ g/cm}^3$  were entered into the fit. For the double layers, Cr-films with a thickness of  $d_{Cr} = 0.42$  to  $0.67 \mu\text{m}$  were deposited as interlayer between the Si substrate and ta-C films with intending to reduce the internal stresses in the hard and stiff carbon layers. Fig. 14 reveals that the Cr interlayer significantly changes the shape of the dispersion curve compared to the sample with the single ta-C layer. In addition to Young's modulus  $E_{ta-C}$ , the density  $\rho_{ta-C}$  and the thickness  $d_{ta-C}$  of the ta-C films, the thickness  $d_{Cr}$  of the Cr interlayer could also be calculated by the fit procedure. These results demonstrate that the surface acoustic wave method is also able to obtain material properties from beneath the surface. For the fitting procedure, the transfer matrix model was used (Thomson 1950; Haskell 1953).

## Applications

### Super-Hard Tetrahedral Amorphous Carbon (ta-C)

Surface films of tetrahedral amorphous carbon ta-C are increasingly used for wear protection. With a high content of the tetrahedral  $sp^3$ -bonds (diamond bonds), these coatings can achieve super-hardness (Vickers hardness  $H > 4000$  HV) (Schultrich et al. 1996; Silva et al. 1996; Robertson 2002). These diamond-like carbon films are an example for a chemical element whose properties vary within a wide range from soft such as graphite ( $sp^2$ -bonds) to hard such as diamond ( $sp^3$ -bonds), depending on the bonding condition determined by the deposition technology. ta-C films are usually produced by physical vapor deposition (PVD) technologies which can generate a plasma with high energy (Schuelke et al. 1999). Measuring the elastic modulus by laser-induced surface acoustic waves has proven a useful method for fast and reliable evaluation of the quality of these super-hard films.

Figure 15a, b shows the correlation of Young's modulus  $E$  of ta-C films with both the content of  $sp^3$  bonds  $S_C$  and the hardness  $H$ . The  $sp^3$  content of ta-C films was detected by means of the x-ray photoelectron spectroscopy (Retzko and Unger 2003). Young's modulus of the films was measured with the laser-acoustic device in Fig. 5. The films had a thickness between 50 and 120 nm. The film properties were varied by increasing the substrate temperature in the process of laser-arc deposition (Schneider et al. 1998b). High  $sp^3$ -content and high Young modulus are achieved at low deposition temperature. Increasing the temperature reduces both  $sp^3$  content and

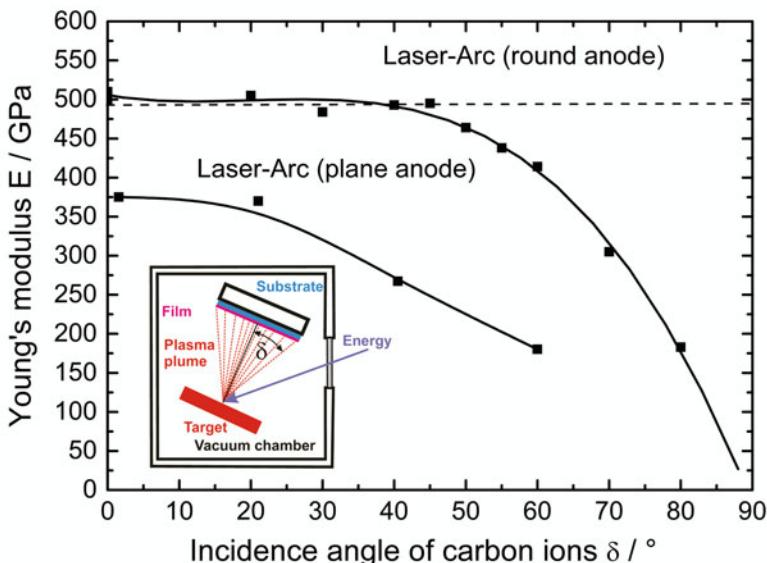


**Fig. 15** (a) Correlation between Young's modulus  $E$  and  $sp^3$  content  $S$  for ta-C films. (b) Correlation between Young's modulus  $E$  and hardness  $H$  for ta-C films

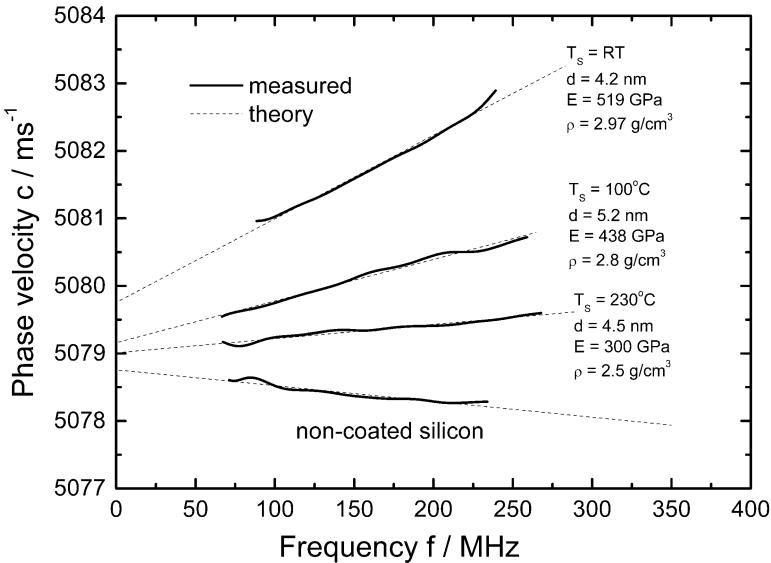
Young's modulus. Figure 15b confirms that the hardness  $H$  correlates with Young's modulus  $E$ , because both,  $H$  and  $E$ , depend on the stiffness of the inter-atomic bonds for this material, ta-C. The correlation between hardness and Young's modulus,  $H = E/10$ , was obtained for films with a thickness  $d > 1 \mu\text{m}$ , which enabled the hardness to be measured accurately.

Figure 16 shows Young's modulus  $E$  versus the angle of incidence of the plasma plume  $\delta$  (Schulz 2005). In the deposition process of ta-C, usually high energy is directed to a target of graphite which generates a plasma whose ionized particles bombard the surface of a substrate where they are deposited as films. The films were deposited with the laser-arc technology on a large silicon wafer. Figure 16 reveals the homogeneity of the deposition conditions for two different technical versions of the electrode igniting the plasma. This is an example that the laser-acoustic method is helpful for improving and optimizing the technology.

Figure 17 shows dispersion curves of ta-C films with a thickness of only 4.2 to 5.2 nm deposited on (100) silicon (Schneider et al. 2002a). Such films were developed with the intention to protect computer hard disks. The films were deposited by high current pulsed vacuum arc evaporator (HCA) at different substrate temperatures  $T_s$  (Schuelke et al. 1997). The dispersion is very small, a velocity difference of about 1 m/s in the frequency range of 100 MHz. Detecting the effect of these ultrathin films requires measurement of the velocity with an error of  $\Delta c \leq \pm 0.1 \text{ m/s}$ . Increasing the measuring distance to 25 mm and enhancing the signal-to-noise ratio by averaging the signals more than 50 times reduce the



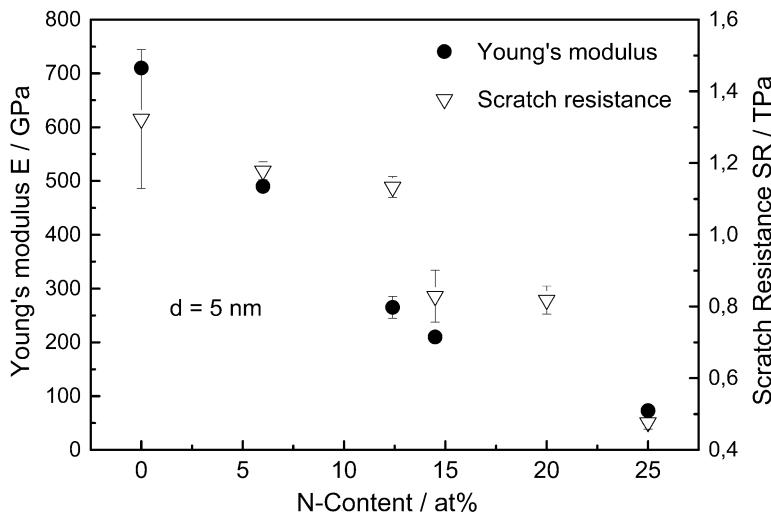
**Fig. 16** Young's modulus  $E$  of ta-C films deposited with laser-arc depending on the angle of incidence  $\delta$  of the plasma plume for two electrode configurations (Schulz 2005)



**Fig. 17** Dispersion curves of ta-C nanometer films deposited with high current pulsed vacuum arc on (100) silicon at different substrate temperature  $T_s$ , Young's modulus  $E$ , density  $\rho$ , and the thickness  $d$  of the films are given (Schneider et al. 2002, reprinted with permission of Elsevier, original Fig. 5)

uncertainty of the measurement. The dispersion curves were fitted with a number of data points  $N > 1000$ . The error of the fit reduces with  $1/\sqrt{N}$  so that the fit has the effect of an additionally smoothing of the dispersion curve. Repeating measurements resulted in a scatter of less than 10% for Young's modulus  $E$  calculated for the films, even for the thinnest film with  $d = 4.2 \text{ nm}$ .

The slope of the curves reveals the effect of the films. Only one film parameter, the Young's modulus  $E$ , could be fitted for these curves. The film thickness  $d$  was measured by an ellipsometer, and the density  $\rho$  was again obtained from the empirical correlation between Young's modulus and the density in Schneider et al. (1997). For all films, the Poisson's ratio was set to  $\nu = 0.19$ . It is to note that already the non-coated substrate had a measurable dispersion, but with negative slope. Having assumed an adsorption layer of  $\text{SiO}_2$  with the tabulated values of  $E = 69 \text{ GPa}$ ,  $\nu = 0.17$ , and  $\rho = 2.2 \text{ g/cm}^3$ , a thickness of  $d = 2.4 \text{ nm}$  was calculated. For these ultrathin films, the  $\text{SiO}_2$  layer should not be neglected in the fit procedure. The values given in Fig. 17 confirm also for the ultrathin ta-C films that the lower the deposition temperature  $T_s$  is, the higher Young's modulus  $E$ . The film with the highest modulus  $E = 519 \text{ GPa}$  was deposited at room temperature  $T_s = \text{RT}$ . Considering Fig. 15a and b, this film is expected to have the highest  $\text{sp}^3$ -content and also the highest hardness.



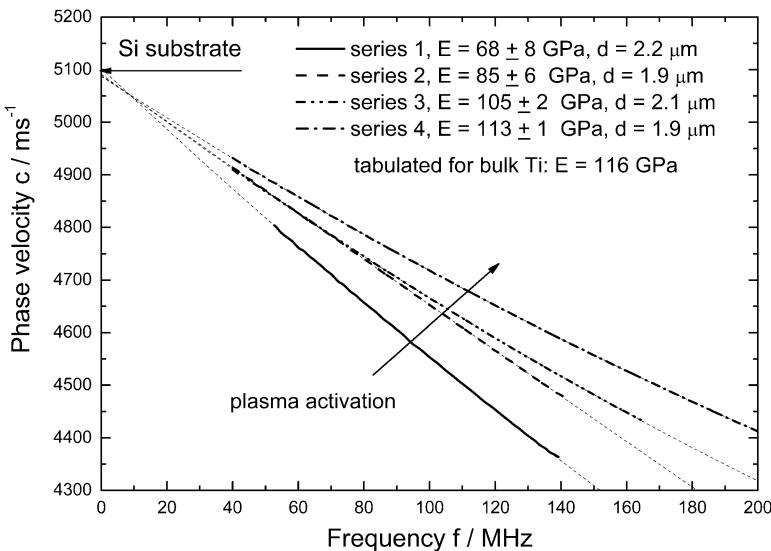
**Fig. 18** Scratch resistance  $SR$  and Young's modulus  $E$  of carbon films with different nitrogen content, deposited by high current pulsed vacuum arc (Schneider et al. 2002, reprinted with permission of Elsevier, original Fig. 8)

Measuring the hardness of nanometer films is difficult. A sensitive method for quantitatively evaluating the plastic deformability of the near surface region on the nanometer scale is a nano-scratch method developed by Wiess et al. (1999), which is based on an AFM equipped with a diamond-tipped cantilever. The method yields the scratch resistance which is calculated from the ratio of the normal force and the cross-sectional area of the residual scratch. Figure 18 shows the Young's modulus  $E$  and the scratch resistance  $SR$  for ta-C films with a thickness of about  $d = 5$  nm and varying nitrogen content. Both the Young's modulus and the scratch resistance reduce with increasing nitrogen content, which confirms the correlation of Young's modulus and hardness also for ultrathin films.

## Porous Titanium Films

The higher the deposition rate, the higher the probability of micro- and nano-pores in the films. Therefore, the deposition processes has to be carefully optimized to deposit dense films at acceptable costs. Pores are micro-defects reducing the strength and the adhesion of the film. Pores may form paths for aggressive media to the substrate, reducing the protective effect of the film against corrosion attack. For wear and corrosion protection, dense films are desired.

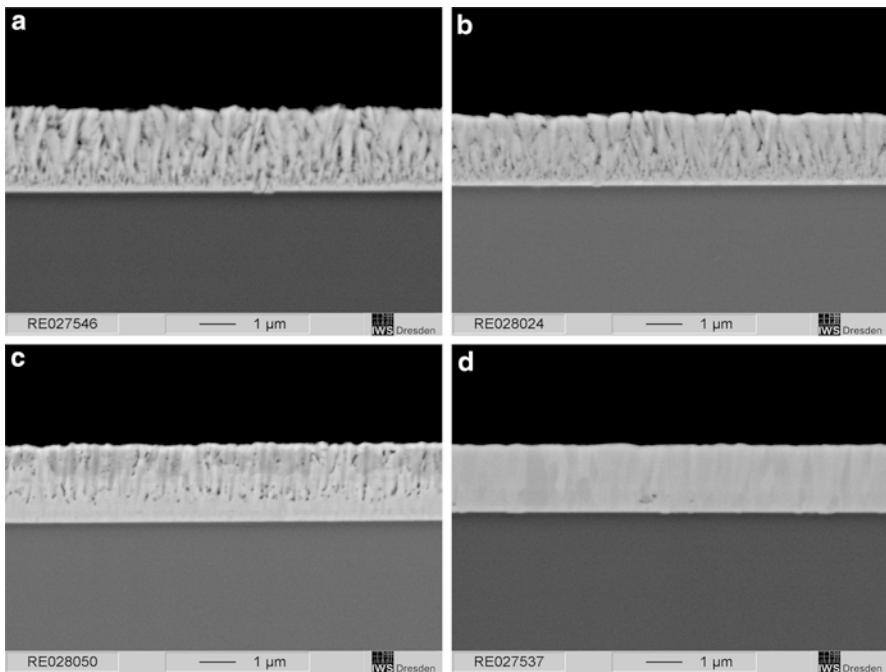
Figure 19 shows dispersion curves measured for four samples of titanium films deposited on silicon (Leonhardt et al. 2004). The films had a thickness between 1.9 and 2.2  $\mu\text{m}$ . They were deposited with the electron beam evaporation process and



**Fig. 19** Dispersion curves measured on (100) silicon coated with titanium films deposited with different plasma activation (Leonhardt et al. 2002, reprinted with permission of Elsevier, original Fig. 7)

DC-vacuum arc discharge. The electrode arrangement of the DC-vacuum arc discharge was successively improved to enhance the activation of the plasma and therewith the effect of the bias voltage to densify the film material. The laser-acoustic test served as an indicator for the porosity of the films.

The dispersion curves in Fig. 19 decrease, since Young's modulus of titanium ( $E_{Ti} = 116 \text{ GPa}$ ) is lower compared to the silicon substrate ( $C_{II} = 165 \text{ GPa}$ ) and the density  $\rho_{Ti} = 4.51 \text{ g/cm}^3$  is higher than  $\rho_{Si} = 2.33 \text{ g/cm}^3$ . Figure 19 also shows the values of Young's modulus  $E$  calculated for the films, representing the mean value of five samples in each test series. The curves allow only one film parameter, Young's modulus  $E$ , to be fitted with acceptable low uncertainty. Therefore, the film density was kept constant at the tabulated value  $\rho_{Ti} = 4.51 \text{ g/cm}^3$  and the film thickness was measured by profilometry. Considering Eq. 1, Young's modulus  $E$  is proportional to  $E \sim c^2 \rho$ , with  $c$  denoting the sound velocity measured and the density  $\rho$ . The tabulated density value  $\rho_{Ti}$  is supposed to be the upper limit for titanium. Therefore, the values for Young's modulus  $E$  in Fig. 19 are the highest possible values. The results for Young's modulus  $E$  reveal that the films of series 1, deposited with electron beam evaporation without plasma activation, have a value of only  $E = 68 \text{ GPa}$  which is 40% lower than the value of bulk titanium. Increasing the activation of the plasma in the film deposition process increases the Young modulus of the film up to the value  $E = 113 \text{ GPa}$  which is close to that of the bulk titanium,  $E_{Ti} = 116 \text{ GPa}$ . Since only pure metal was deposited, the low values of the elastic modulus must be the result of porosity in the films.



**Fig. 20** (a) SEM picture of a Ti film of series 1,  $E = 68$  GPa. (b) SEM picture of a Ti film of series 2,  $E = 85$  GPa. (c) SEM picture of a Ti film of series 3,  $E = 105$  GPa. (d) SEM picture of a Ti film of series 4,  $E = 113$  GPa (Leonhardt et al. 2002, reprinted with permission of Elsevier, original Fig. 8a–d)

Figure 20a–d show micrographs of the cross sections of four samples, made with a scanning electron microscope JSM 6400. The micrographs are presented in the order of increasing film modulus. The technology of electron beam evaporation can achieve a high deposition rates ( $>10$   $\mu\text{m}/\text{min}$ ) on large dimensions. However, the micrograph in Fig. 20a confirms for test series 1 that the low particle energy in the vapor produced titanium films with high porosity and low compactness. The micrograph from test series 2 (Fig. 20b) shows also a distinct porosity with a similar character as in series 1. However, the total porosity is obviously lower, correlating with the higher modulus of  $E = 85$  GPa measured for this sample series. The pores in the sample of test series 3 (Fig. 20c) are small and globular. The diameters of the pores are not larger than 50 nm. Although the laser-acoustic method cannot detect single pores, the volume effect of all of them together is indicated by a Young's modulus of  $E = 105$  GPa that is still lower than the value of the bulk material ( $E_{Ti} = 116$  GPa). The films of test series 4 (Fig. 20d) with the highest film modulus of  $E = 113$  GPa are nearly completely dense, apart from some isolated defects that were found very seldom in these films. Within the error of the laser-acoustic measurement, the Young's modulus measured for this film agrees with the bulk value tabulated for titanium  $E_{Ti} = 116$  GPa. Comparing

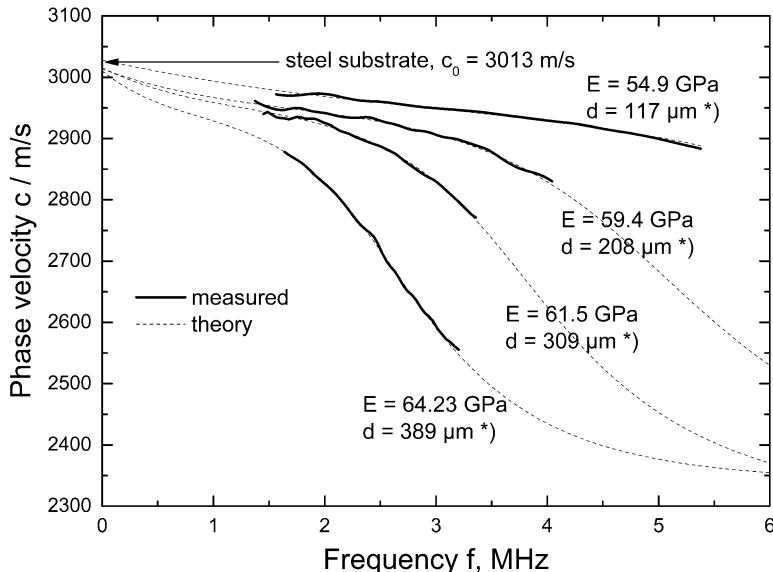
the test series 1 to 4 reveals that combining the process of electron beam evaporation with the process of DC-vacuum arc discharge whose plasma-activation was improved step by step enabled a deposition technology to be developed which produces dense titanium films with acceptable efficiency. The method of laser-induced surface acoustic waves was helpful in testing the film quality.

There are also technological developments which are aimed to deposit films with high porosity. This applies for low-k silica xerogel SiCOH-films which reduce the dielectric permittivity in interconnect lines of semiconductor circuits. A porosity of up to 50% is strived for these films. This very high porosity leads to low mechanical stability of these films. This is an object of ongoing research which also uses laser-induced surface acoustic wave for the material characterization (Flannery et al. 2001; Xiao and You 2006; Schneider et al. 2005; Xiao et al. 2016).

## Thermal-Sprayed Coatings

The technologies of thermal spraying can make coatings with a wide variety of properties. Although modern processes, which can spray nearly all kinds of materials such as metals, ceramics, or hard metals, can deposit nearly dense coatings, residual porosity and small cracks cannot always be avoided. Therefore, technological developments and quality control require fast and nondestructive test methods to evaluate the mechanical integrity of the coating materials. Young's modulus can be used as a sensitive indicator for the material integrity, because pores and micro-cracks can reduce the stiffness of the material considerably (Schneider et al. 1993; Lima et al. 2005; Bescond et al. 2007; Berger et al. 2012).

Figure 21 shows dispersion curves measured on  $\text{Al}_2\text{O}_3$ -coatings with a thickness in the range from 117  $\mu\text{m}$  to 389  $\mu\text{m}$  deposited on steel S235JR plates with a dimension of 100  $\times$  60  $\times$  5 mm. The coatings were sprayed with Atmospheric Plasma Spraying (APS). At the zero frequency the fitted curves start all at the velocity  $c_0$  of the steel substrate, about 3013 m/s. All curves decay, indicating the velocity of the  $\text{Al}_2\text{O}_3$ -coatings to be lower than the velocity of the steel substrate. The thicker the  $\text{Al}_2\text{O}_3$  film, the steeper the curve decays. The measurements were done on the as-sprayed surface. The high roughness and the, usually, inhomogeneous microstructure of the coatings cause high attenuation of the wave. As a consequence, surface acoustic waves were only detected in the frequency range  $f < 6$  MHz. For this application, a commercially available wedge-transducer with a frequency range from 1 to 20 MHz was used. Fitting the curve yields values between  $E = 54,9$  and  $64,23$  GPa for Young's modulus of the  $\text{Al}_2\text{O}_3$  coatings. Only Young's modulus could be calculated for the coatings. The coating thickness was measured by a Dualscope MP200 with an EGA2H sensor (Helmut Fischer GmbH, Germany). The coating density was determined by the gravimetric method on two samples and was found to be  $\rho = 3.97$  g/cm<sup>3</sup>. These values were entered into the fitting procedure. Apart from APS (Atmospheric Plasma Spraying: F6 torch, GTV, Germany), two other technologies were used to spray

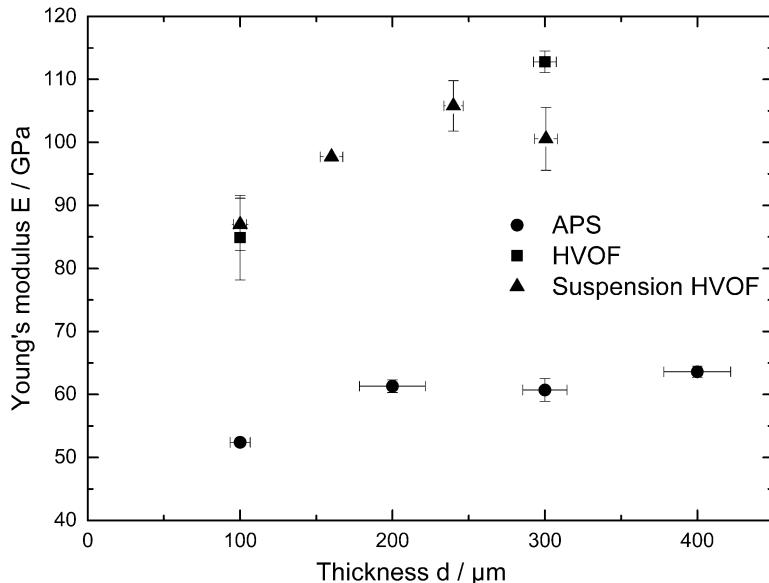


**Fig. 21** Dispersion curves of surface acoustic wave measured on  $\text{Al}_2\text{O}_3$  coatings deposited with Atmospheric Plasma Spraying (APS) on steel, \*) thickness  $d$  entered into the fit (Berger 2011)

$\text{Al}_2\text{O}_3$ -coatings, HVOF (High-Velocity Oxy Fuel: Top Gun, GTV, Germany) and Suspension HVOF (Berger 2011).

Figure 22 shows Young's modulus  $E$  versus the coating thickness  $d$  for  $\text{Al}_2\text{O}_3$  coatings. The surfaces of the test sample were divided into ten sections. Young's modulus was measured with the laser-acoustic method three times to get a mean value and standard deviation. The data presented in Fig. 22 are the mean values of all ten sections of each sample. The APS coatings have the lowest Young's modulus  $E$ , about 60 GPa. It is considerably lower than the Young's modulus of  $E = 350$  GPa known for bulk  $\text{Al}_2\text{O}_3$  (Asmani et al. 2001). HVOF- and Suspension HVOF-sprayed coatings have Young's modulus  $E$  between 87 and 112 GPa, which is up to twice higher than the values of the APS coatings, indicating these coatings to be more compact. Furthermore, the modulus of the HVOF and the Suspension HVOF coatings distinctly increases with the thickness, suggesting that the defect structure in the films has a gradient perpendicular to the surface with more coherent material in the region nearer to the surface.

The question arises, if Young's modulus can provide more information about the microstructure of the material. The theory of heterogeneous materials is a suitable approach to estimate the effect of some characteristic properties of the porosity on the Young's modulus. The basic idea of this theory is that, from a macroscopic point of view, the elastic behavior of a complex real material with many pores of different shapes and dimensions can be characterized by an effective modulus of a homogeneous equivalent material. It assumes the microstructural elements to be small compared to a characteristic length so that the effect of all microstructural elements

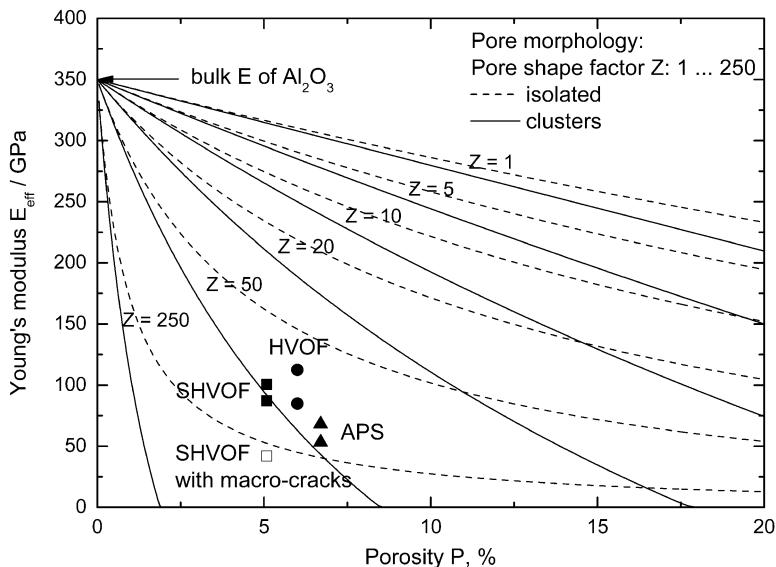


**Fig. 22** Young's modulus E versus thickness d of  $\text{Al}_2\text{O}_3$  coatings deposited by thermal spraying with APS, HVOF, and Suspension HVOF (Berger 2011)

can be averaged for a volume with the dimensions of this characteristic length. In the present case, the characteristic length is the wavelength  $\lambda$  of the surface acoustic wave.

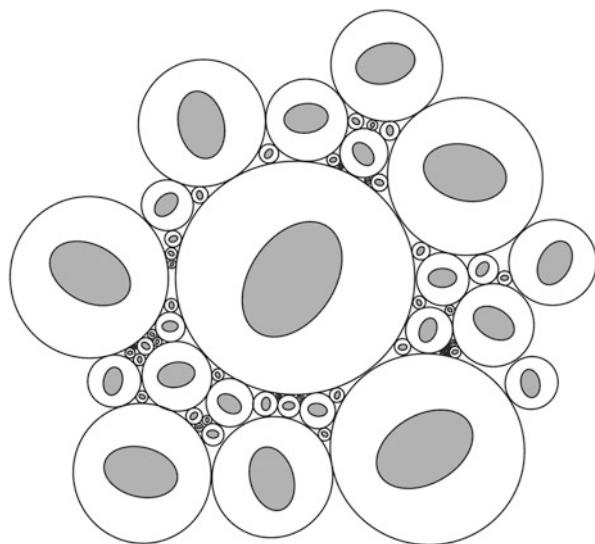
Figure 23 presents two sets of theoretical curves showing the dependence of Young's modulus of  $\text{Al}_2\text{O}_3$   $E_{\text{eff}}$  on the porosity  $P$ . They were calculated with a statistical continuum mechanics model proposed by Kreher and Pompe (1989) for non-dilute porosity. The theory for the effective modulus of a stochastic material structure cannot be developed without substantial simplifications. The following basic assumption is made: the pore is an ellipsoid whose shape can be described by the ratio of the major and minor semi-axes ( $a$  and  $b$ ) termed the pore shape factor  $Z = a/b$ . The problem of an ellipsoidal inclusion in a homogeneous matrix was solved by Eshelby (1957). The pore is isolated but interacts with other pores. The interaction is taken into account by a self-consistent method termed the effective field approach (Kröner 1961; Hill 1965). The homogeneous matrix, in which the pore is embedded, is assumed to be influenced by an averaged effect of other pores. The pores are supposed to be arranged according to the composite sphere model (Hashin 1962). The basic elements of this arrangement are spheres consisting of ellipsoidal inclusions embedded in the equivalent matrix (see Fig. 24).

The largest of them are small compared to the characteristic dimension of the body, in the present case, the wavelength  $\lambda$ . The ratio of the volume of the inclusion to the volume of the sphere is equated to the volume fraction of the inclusion and



**Fig. 23** Effective Young's modulus  $E_{\text{eff}}$  versus porosity  $P$  for  $\text{Al}_2\text{O}_3$  and the values measured for  $\text{Al}_2\text{O}_3$  coatings sprayed with APS, HVOF, and Suspension HVOF (Berger 2011)

**Fig. 24** Arrangement of the elements in the composite sphere model according to the theory of the elastic modulus of heterogeneous materials (Hashin 1962)



does not depend on the sphere diameter. The gaps between the larger spheres are filled with smaller spheres with the same composition until the body volume is completely continuous. This arrangement assumes a wide spectrum of pore dimensions.

The ellipsoids are randomly oriented. For the general case of solid inclusions, the problem has to be solved numerically. For pores, a simplified approximation can be given (Kreher and Janssen 1992) as.

$$E_{\text{eff}} = E_0 \frac{1 - P}{1 + P(T_{\text{eff}} - 1)} \quad (42)$$

$E_{\text{eff}}$  denotes the effective modulus of the material containing pores.  $E_0$  is Young's modulus of the matrix material which is free of pores. The parameter  $T_{\text{eff}}$  is a function depending on the pore shape  $Z$  and the Poisson's ratio  $\nu$  of the matrix material.

The curves in Fig. 23 were calculated for six pore shapes  $Z = 1, 5, 10, 20, 50$  and  $250$ . A Young's modulus of  $E_0 = 350$  GPa was assumed for bulk  $\text{Al}_2\text{O}_3$  (Asmani et al. 2001). A pore shape factor of  $Z = 1$  denotes a spherical pore, whereas  $Z > 1$  describes disk-like pores of increasing flatness up to cracks.

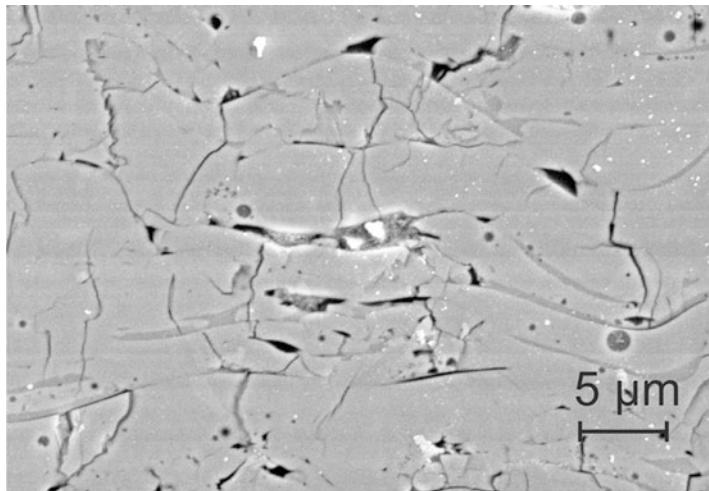
For the first set of curves, the pores were assumed to be isolated and surrounded by the  $\text{Al}_2\text{O}_3$  matrix. Spherical pores ( $Z = 1$ ) reduce the effective elastic modulus in a moderate way, nearly proportional to the fraction of pore volume in the material. In contrast, disk-like pores have a considerably higher effect. The flatter the disk, the more the elastic modulus decreases. As the porosity approaches  $P = 100\%$ , the effective modulus approaches zero.

The second set of curves was calculated assuming the pores to form random clusters. With increasing pore shape factor  $Z$ , the porosity has a dramatic effect on the effective modulus  $E_{\text{eff}}$ . The curves of  $E_{\text{eff}}$  descend very steeply and can reach the value  $E_{\text{eff}} = 0$  at a porosity considerably lower than  $P = 100\%$ . For a pore shape factor of  $Z = 50$ , the material can disintegrate at a porosity of  $P = 8.5\%$ .

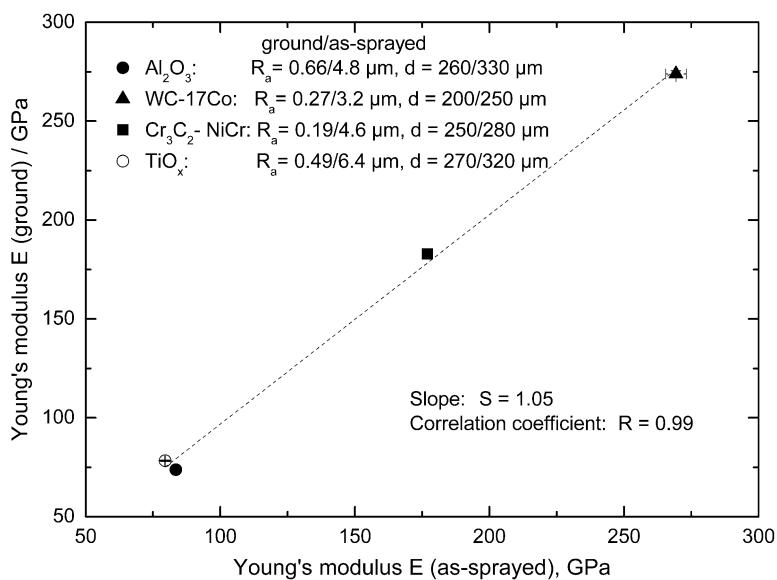
The diagram in Fig. 23 also shows the data points measured for the  $\text{Al}_2\text{O}_3$  coatings sprayed with APS, HVOF, and Suspension HVOF. They reveal that the pore morphology in all three cases deviates strongly from the spherical shape. Flat crack-like pores of high aspect ratio  $Z$  can be expected. They may also form clusters penetrating some regions of the coating material. Figure 25 presents a micrograph of a high-resolution scanning electron microscope. It illustrates that the  $\text{Al}_2\text{O}_3$ -coatings were not formed as a continuous material, but have many micro-cracks reducing the mechanical cohesion. The micrograph of the microstructure confirms the conclusion drawn from the measurement of Young's modulus. Similar observations were made for thermal-sprayed  $\text{ZrO}_2$  coatings (Schneider et al. 1993).

Thermally sprayed coatings may have rough surfaces and the question arises, if this roughness influences the results of the surface acoustic wave method. This effect was investigated for  $\text{Al}_2\text{O}_3$  and  $\text{TiO}_x$  coatings, as well as for WC-17Co and  $\text{Cr}_3\text{C}_2\text{-NiCr}$  coatings sprayed with APS or HVOF, respectively. For each sample of this test series, one half of the surface was carefully ground and the other half remained in the as-sprayed state. Young's modulus was measured by laser-acoustics at least five times in both regions of the coatings. The surface roughness was measured with a Mahr Perthometer M1.

Figure 26 shows Young's modulus measured at the ground surface versus the modulus measured at the as-sprayed surface. The diagram also gives the mean



**Fig. 25** SEM micrographs of an  $\text{Al}_2\text{O}_3$  coating sprayed with Atmospheric Plasma Spraying (APS) (Berger 2011)



**Fig. 26** Comparison of Young's modulus  $E$  measured with surface acoustic waves on as-sprayed and ground surfaces of four different coatings (mean roughness index  $R_a$  and thickness  $d$  are given for both surface conditions) (Berger 2011)

roughness index  $R_a$  (C.L.A.) and the coating thickness  $d$  in the ground and as-sprayed regions. The linear regression line has a slope of  $S = 1.05$  and a regression coefficient of  $R = 0.99$ . These values lead to the conclusion that the roughness did not influence the results of the surface acoustic wave method up to a mean roughness index  $R_a \leq 6.35$ , which is the highest roughness of the investigated samples. This can be ascribed to the low ratio of the roughness  $R_a$  to the wavelength  $\lambda$  of the surface acoustic wave,  $r = R_a/\lambda$ . Assuming the highest frequency  $f = 10$  MHz and the lowest velocity of  $c = 2500$  m/s allow the calculation of the minimum wavelength of  $\lambda = 250$   $\mu\text{m}$ . The ratio of roughness to wavelength was not higher than  $r = 0.025$ .

For a variety of materials, Berger et al. (2012) have compared the Young's modulus of bulk materials and thermal-sprayed coatings which were produced with different spray technologies.

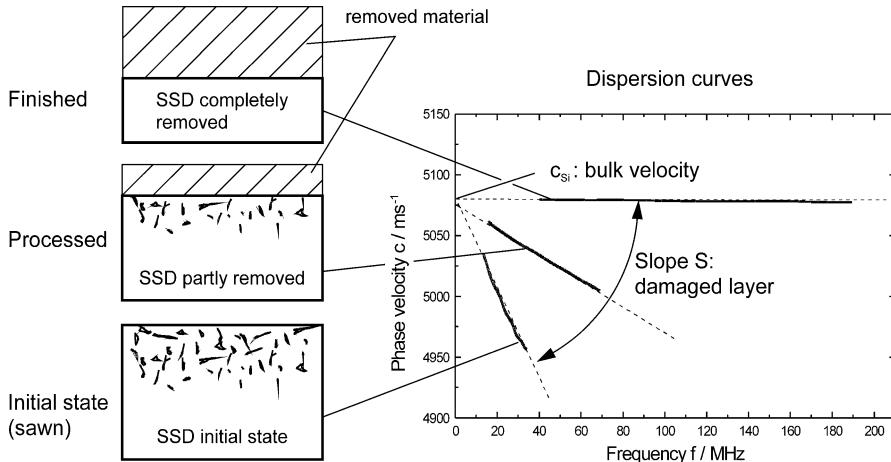
## Subsurface Damage

Cutting materials and machining the surface of technical components by milling, grinding, or polishing damage the material in the region beneath the surface. In metals, this sub-subsurface damage is a region of high plastic deformation, deformation textures, residual stresses, and enhanced defect density.

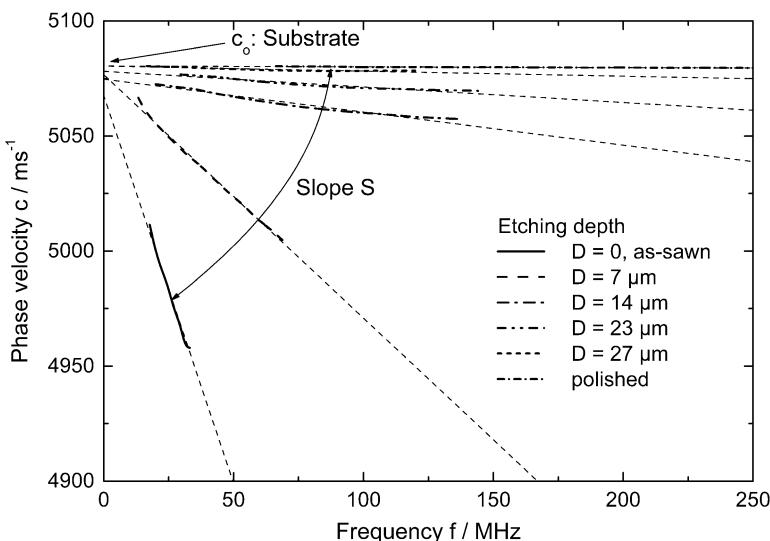
For brittle semiconductor materials, such as silicon or GaAs, cutting a wafer from an ingot produces a network of micro-cracks at the surface, which must be removed completely by careful grinding, etching, and polishing processes. Optimizing the processes of sawing and surface machining is an important factor for low costs, high quality, and high reliability in the wafer manufacturing. It is useful to have reliable methods for characterizing the subsurface damage. Therefore, it was investigated, if surface acoustic waves are suitable for evaluating the subsurface damage (Schneider et al. 1999, 2002b; Paehler et al. 2007).

Figure 27 shows dispersion curves measured on silicon wafers with three different surface states: as-sawn from the ingot, after a part of the subsurface damage (SSD) was removed by etching and grinding, and the final polished state. In the case of the as-sawn state, the curve steeply descends with the frequency, which is typical for a material with a surface layer whose Young's modulus is much lower compared to the substrate. This result confirms that the cutting process had produced a layer of high defect density. Especially, the micro-cracks drastically reduce the elastic modulus in the layer of the subsurface damage. This effect was already described for thermal-sprayed coatings in section “[Thermal-Sprayed Coatings](#).” Removing a part of the subsurface damage reduces the slope of the dispersion curve. For the final polished surface, the dispersion curve is a straight line parallel to the frequency axis which is typical for completely homogenous materials. At zero frequency, all curves converge to the same point, the velocity of the non-damaged substrate. This behavior suggested that the dispersion of the surface wave can be used to determine the depth of the subsurface damage. However, theoretical modeling of the dispersion in a material with subsurface damage is difficult, due to the complex structure of the

damage layer. The damage layer is assumed to consist of three different zones (Paehler et al. 2007). The first zone is some tens of nanometers thick. The material is completely damaged. The second zone is some micro-meters deep. It is heavily damaged with a network of many cracks preferentially oriented perpendicular to the surface. The third zone is slightly damaged with comparably long cracks reaching deeper than 10 µm into the material.



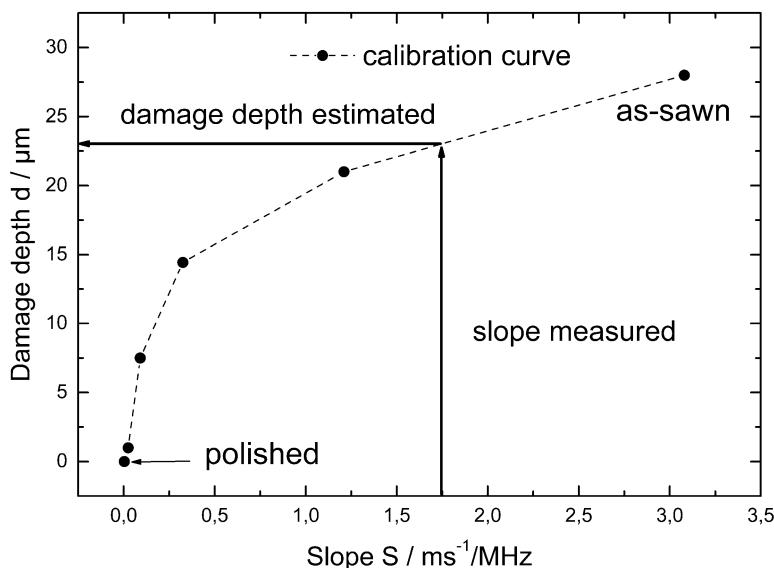
**Fig. 27** Subsurface damage causes dispersion of surface acoustic waves (Paehler et al. 2007, reprinted with permission of Elsevier, original Fig. 4)



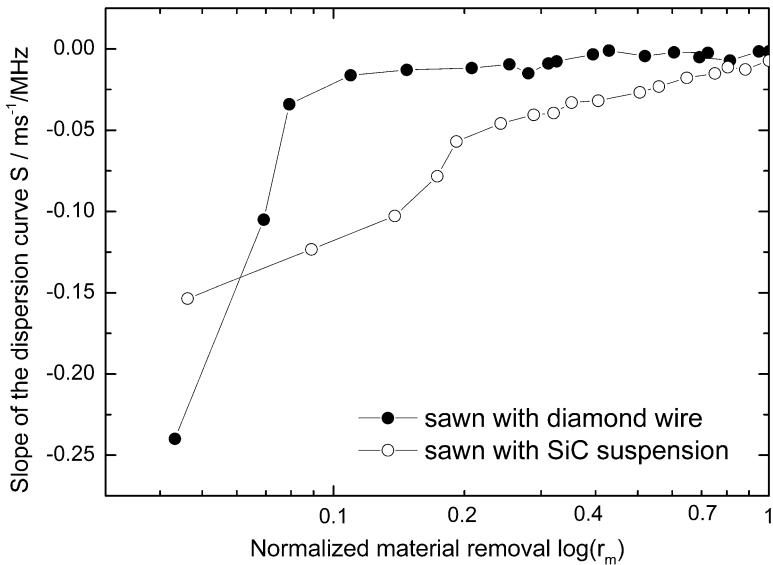
**Fig. 28** Dispersion curves measured on an (100) silicon wafer sawn from the ingot and stepwise etched to remove the subsurface damage (Schneider 2012)

Due to the complexity in modeling, the method must be calibrated with samples whose damage depth is determined with an alternative method. Figure 28 shows the dispersion curves measured on a 300 mm silicon wafer which was stepwise etched up to the depth at which the slope of the dispersion curve does not change anymore. The etching depth  $D$  was obtained from the weight loss of the wafer. For an etching depth of  $D_0 = D = 27 \mu\text{m}$ , the detectable damage layer was completely removed. Subsequently, the damage depth  $d$  can be calculated for each etching step by  $d = D_0 - D$ . The linear regression yields the slope  $S$  of the dispersion curve and additionally the velocity  $c_0$  of the non-damaged substrate. The values of the damage depth  $d$  and the related values of the slope  $S$  allow the set up of a calibration table shown as diagram in Fig. 29. With this calibration table, the damage depth of an unknown test wafer can be determined in three steps: 1. the dispersion curve  $c(f)$  is measured, 2. the slope  $S$  is calculated by linear regression, 3. the value of the damage depth  $d$  related to the slope  $S$  is read from the calibration table by interpolation.

It is to note that the calibration table is not independent on the sawing technology. Figure 30 shows two examples from the solar cell wafer industry (Schneider 2013). The diagram presents the slope  $S$  of the dispersion curve versus the material removal  $r_m$  which is given in normalized values for reasons of confidentiality. Solar cell wafers of (110) silicon single crystals were wire-sawn with two technologies, with diamond wire and with SiC suspension, respectively. The depth profile of the damage is clearly different. The diamond wire sawing produced a higher magnitude of the slope  $S$ . However, removing only a small amount of material considerably reduces the



**Fig. 29** Damage depth  $d$  versus the slope  $S$  of the dispersion curve, used for the evaluation of the subsurface damage (Schneider 2012)



**Fig. 30** Slope  $S$  of the dispersion curve versus normalized removal  $r_m$  for (110) single crystal solar cell wafers sawn with diamond wire and with SiC suspension (Schneider 2013)

negative slope. Already after the first etching step, the slope  $S$  is smaller for the wafer sawn with diamond wire than in the wafer sawn with SiC suspension. After the fourth removal, the slope  $S$  of the dispersion curve is close to the value of the bulk material.

The sawing with SiC suspension produces a lower magnitude of the slope  $S$  than the diamond wire sawing. But the slope approaches more gradually to the zero value. Comparing both profiles leads to the conclusion that the diamond wire sawing produces a damage which is more concentrated near the surface and does not influence the material as deep as the wire-sawing with SiC suspension.

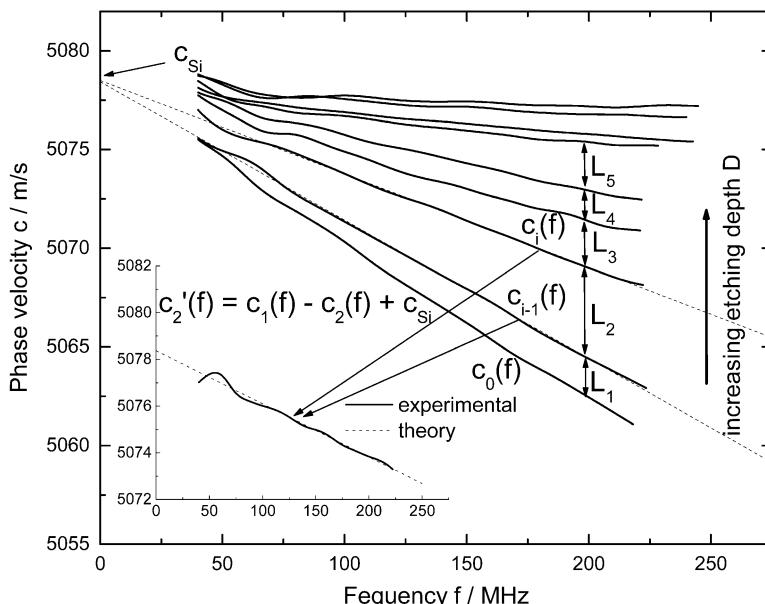
Paeehler et al. (2007) used the surface acoustic wave method to determine profiles of the Young modulus of the subsurface damage in silicon wafers machined by rotational grinding. Since Young's modulus is very sensitive to the density of micro-defects, this profile can analyze the damage gradient in more detail than the simple measurement of the slope  $S$  of the dispersion curve. The essential process was a reactive ion etching technique (RIE) used for a precise stepwise removing of the subsurface damage. Material was removed in steps from 19 nm up to 6000 nm by varying the etching time from 5 s to 1600 s. The thickness of the removed layer was again determined by measuring the weight before and after etching, using a Mettler AT201 high accuracy balance with 10  $\mu\text{g}$  resolution.

Figure 31 shows the dispersion curves  $c_i(f)$  measured in a stepwise etched wafer, with  $i$  denoting the number of the etching step.  $c_0(f)$  denotes the dispersion curve of the non-etched surface. The profile of the damage was approximated by a stack of

layers with different Young's moduli. Within a layer the modulus is assumed to be constant. The effect of a single layer  $L_i$  was isolated by subtracting the dispersion curve  $c_i(f)$  measured after the etching step  $i$  from that one measured before,  $c_{i-1}(f)$ . The difference  $c_{i-1}(f) - c_i(f)$  represents the effect of layer  $L_i$  on the dispersion. Adding the wave velocity  $c_{Si}$  of the non-damaged silicon according to Eq. 43 results in a dispersion curve which can be analyzed by the fitting procedure with the theory in section “[Surface Acoustic Waves](#)” to calculate Young's modulus  $E_i$  of the layer  $L_i$ .

$$c'_i(f) = c_{i-1}(f) - c_i(f) + c_{Si} \quad (43)$$

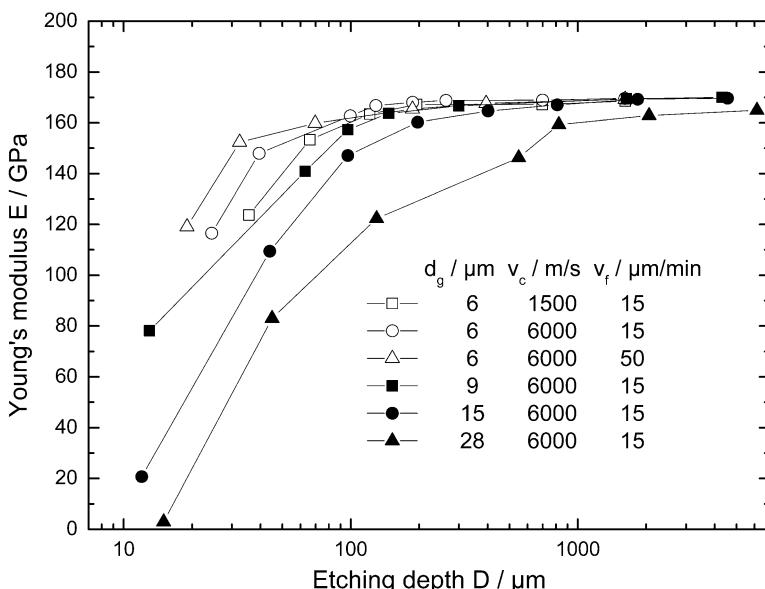
The value of  $c_{Si}$  was determined by extrapolating the dispersion curves to  $f = 0$  MHz, as shown in Fig. 31. For the example of layer  $L_2$ , Fig. 31 demonstrates how the related dispersion curve  $c'_2(f)$  is derived from  $c_2(f)$  and  $c_1(f)$ , using relation (43). Fitting the differential dispersion curve  $c'_2(f)$  yielded the Young modulus  $E_2 = 152$  GPa for layer  $L_2$ . The differential dispersion curves  $c'_i(f)$  are straight lines. They can provide only two material parameters,  $C_{11}$  of the substrate and Young's modulus  $E_i$  of the layer. The density  $\rho_i$  and the thickness  $d_i$  of the layer  $L_i$  have to be known to calculate  $E_i$ . The layer thickness  $d_i$  was determined from the weight loss after etching step  $i$ . For the density  $\rho_i$ , the value of the non-damaged silicon  $2.33 \text{ g/cm}^3$  was applied. For layers with



**Fig. 31** Dispersion curves  $c_i(f)$  measured for a silicon wafer after increasing etching steps  $i$  and the derivation of the differential dispersion curve  $c_2(f)$  for the layer  $L_2$  removed in etching step 2 (Paehter et al. 2007, reprinted with permission of Elsevier, original Fig. 7)

many defects, the true density is probably lower. The uncertainty arising from this simplification can be estimated, using again Eq. 1. The Young modulus  $E$  is proportional  $E \sim c^2 \rho$ , with  $c$  denoting the sound velocity measured and  $\rho$  the density. Since for the density  $\rho$ , the highest possible value of the non-damaged silicon has been used for all layers  $L_i$ , the values calculated for  $E_i$  represent the upper limits for elastic modulus. Therefore, the shape of the profiles of the elastic properties in the subsurface damage is not false.

Figure 32 shows the profiles of Young's modulus  $E$  measured at wafers machined by rotational grinding with different parameters, average grinding grain size  $d_g$ , cutting speed  $v_c$ , and feed rate  $v_f$ . In all profiles the modulus  $E$  is lowest in the uppermost layer. With increasing depth, the modulus approaches the limit value of  $E \approx 169$  GPa which is Young's modulus of the perfect silicon. Comparing the profiles reveals the distinct influence of grit size  $d_g$  on the degree of subsurface damage. With increasing grain diameter  $d_g$  the modulus  $E_1$  of the uppermost layer  $L_1$  is reduced to lower values and the damage reaches deeper into the material. In the case of  $d_g = 28 \mu\text{m}$ , the modulus  $E_1$  is even lower than 10 GPa. This low value indicates an approximate disintegration and disruption of the crystal lattice within a layer of  $d = 19 \text{ nm}$  from the surface. The damage penetrates deep into the material and the limit value of  $E \approx 169$  GPa is not reached even at the highest material removal of  $D = 6000 \text{ nm}$ . The smallest damage was produced with grit size  $d_g = 6 \mu\text{m}$  and a feed rate of  $v_f = 50 \mu\text{m}/\text{min}$ . Young's modulus near the surface was lowered



**Fig. 32** Profiles of Young's modulus  $E$  versus the etching depth  $D$  for silicon wafers rotationally ground,  $d_g$  average grinding grain size,  $v_c$  cutting speed,  $v_f$  feed rate (Paebler et al. 2007, reprinted with permission of Elsevier, original Fig. 15)

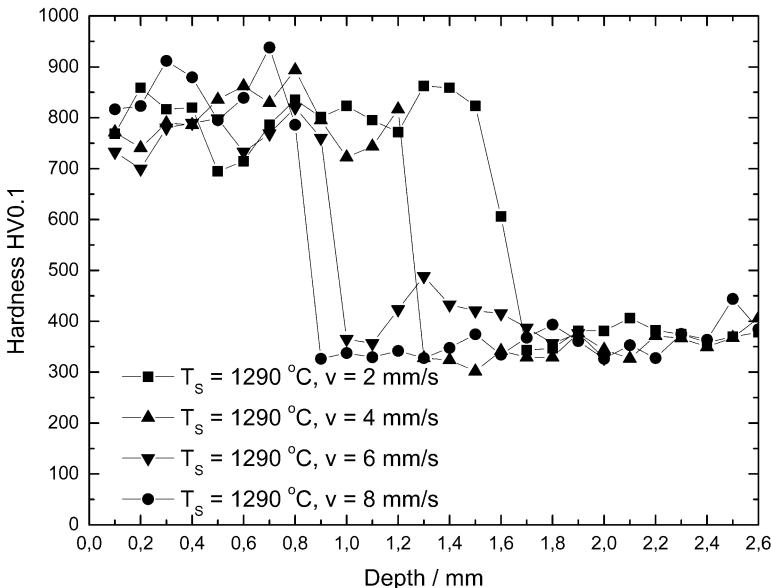
only moderately to  $E \approx 120$  GPa and the bulk modulus is reached at the depth of  $D = 200$  nm. The profiles give a good insight into the structure of the damage and allow to conclude to the damage mechanisms during the grinding process.

## Laser-Hardened Steels

Surface hardening of steel is a widespread technology to combine two essential, but contrary properties of steel in one component; sufficient fracture toughness of the core and high hardness at the surface to improve wear resistance and fatigue strength. There is a variety of technologies available for surface hardening: induction hardening, flame hardening, laser hardening, or case hardening. Surface hardness and hardening depth are essential parameters determining the effect of the hardening on the lifetime of the component. The use of nondestructive test methods is desirable to evaluate the surface hardening structures especially for large and expensive components which cannot be allowed to fail. The techniques of ultrasonic backscattering (Willems 1991) and Barkhausen noise (Dobmann et al. 1988) are known for measuring the hardening depth nondestructively. These methods work well, but only if certain conditions are met, which depend on the range of the hardening depth and the microstructure of the hardened layer and the substrate.

Figure 33 shows hardness profiles measured at metallographic cross-sections of the laser-hardened steel 42CrMo4. The hardening depth was varied by varying the traverse speed  $v$  of the laser at a constant surface temperature  $T_s$  which was controlled by controlling the laser power. The hardening was done by using a robot-controlled laser hardening machine equipped with a 3 kW high-power diode laser Laserline LDL 160–3000, a Reis SRV40 robot, and a thermal camera for controlling the surface temperature with an accuracy of 5 °K (ALOtec Dresden GmbH). The laser hardening is used for local and warp-free hardening of functional surface areas of large components. The hardening depth is typically in the range between 0.8 and 1.5 mm, which is difficult to detect with the known test methods.

Figure 34 presents dispersion curves of the surface acoustic wave measured at the 42CrMo4 samples whose hardness profiles are shown in Fig. 33. These curves were measured with a laser-acoustic handheld probe (Schneider et al. 2012b). The signal processing uses a similar procedure described in section “[Experimental Setup and Measuring Procedure](#)” for the experimental setup in Fig. 5. The laser-acoustic probe was equipped with a 10 MHz sensor. The curves are the result of averaging ten single measurements. The velocity values in the curve had a standard deviation of  $\Delta c \leq \pm 2$  m/s. The curves reveal the effect of laser-hardened layers with increasing depth on the surface wave velocity. The non-hardened sample shows a weak dispersion attributed to the effect of a deformation layer. This layer is the result of sawing the sample and machining the surface. It has an enhanced density of dislocations and the grains have a preferential orientation depending on the machining direction (Schneider et al. 1986). For the laser-hardened samples, the deeper the hardening depth, the more the phase velocity  $c$  decreases with frequency  $f$ . For

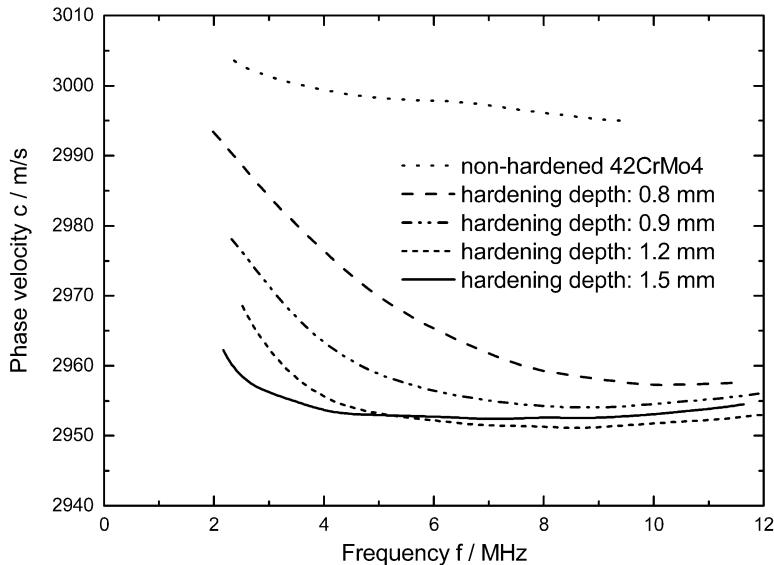


**Fig. 33** Hardness profiles of laser-hardened traces hardened in the steel 42CrMo4,  $T_S$  surface temperature,  $v$  traverse speed (Schneider et al. 2012b, reprinted with permission of Elsevier, original Fig. 2)

frequencies high enough ( $f \geq 9$  MHz), the velocity  $c$  approaches a nearly identical and constant value which depends on the microstructure of the laser-hardened layer. For frequencies  $f \leq 4$  MHz, the velocity  $c$  distinctly correlates with the hardening depth. The metal-physical phenomenon that the velocity of the surface acoustic waves propagating in the martensitic hardened layer is lower than in the non-hardened steel is based on the fact that the body-centered cell of the iron lattice is tetragonally expanded by the carbon atoms trapped interstitially as a result of the martensitic transformation caused by rapid quenching from above the austenitization temperature. This reduces the interatomic forces and consequently Young's modulus  $E$  of the steel (Kurdjumov 1960). The elastic modulus  $E$  and density  $\rho$  determine the acoustic wave velocity  $c$  by the relation  $c \sim \sqrt{E/\rho}$  (Eq. 1). It is important to mention that the density  $\rho$  is also reduced by the martensitic transformation compared to the equilibrium state of iron, Ferrite.

The dispersion curves in Fig. 34 can provide two types of information about the hardened layer, the hardening depth at low frequencies  $f \leq 4$  MHz and the hardness at high frequencies  $f \geq 9$  MHz.

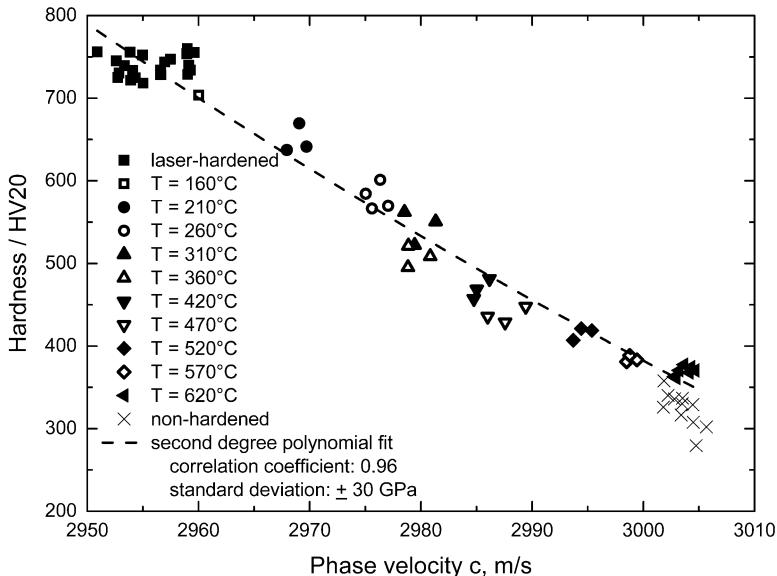
In order to investigate the correlation of the hardness HV with the propagation of the surface acoustic wave  $c$ , laser-hardened samples of steel 42CrMo4 were annealed. The heat treatment was performed in an Argon atmosphere. The samples were inserted into the preheated oven and tempered for 1 h. Afterwards, they were



**Fig. 34** Dispersion curves measured at the laser-hardened 42CrMo4 samples; related hardness profiles are shown in Fig. 33 (Schneider et al. 2012b, reprinted with permission of Elsevier, original Fig. 8)

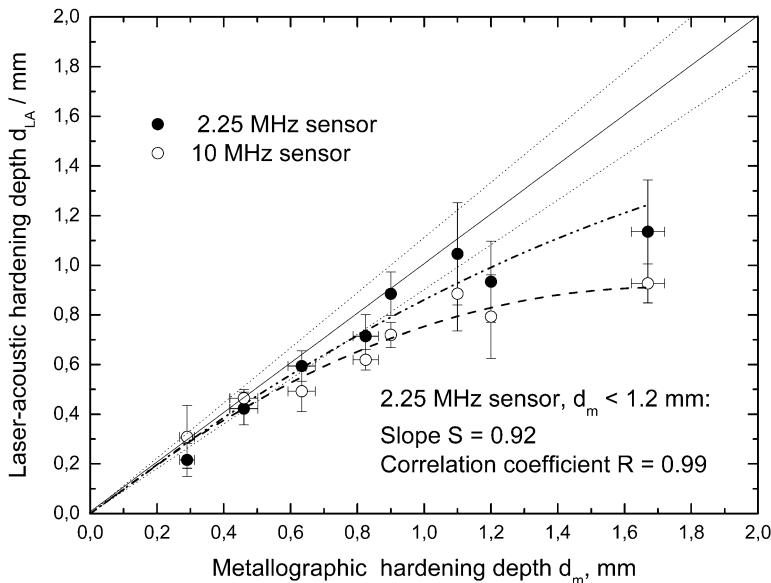
cooled down at room temperature in ambient air. The annealing was started at  $T = 160$  °C and increased in steps of 50 °K. The hardening depth  $d$  of the samples was about  $d = 1$  mm to make sure that the wave at  $f = 9$  MHz propagates completely within the hardened layer and is not influenced by the substrate. The hardness was measured at the surface with a hardness tester Mitutoyo AVK-C1. With an indenter load of 20 N, the test at the relatively rough surfaces could be done with an uncertainty of at least  $\pm 5\%$ .

Figure 35 demonstrates that the hardness HV20 correlates with the velocity  $c$  of the acoustic wave for the annealed martensitic hardened microstructure of the steels. The measurements were done at three different laser-hardened traces. The velocity  $c$  increases, if the hardness HV20 decreases with increasing annealing temperature  $T$ . A similar correlation between hardness HV20 and velocity  $c$  of the surface acoustic wave was demonstrated for the steels C45, X153CrMoV12.1 and 90MnCrV8 (Schneider et al. 2012b). This behavior is ascribed to the relaxation of the martensitic lattice. Annealing relieves the stress in the lattice due to the precipitation of carbon atoms as iron carbide. The hardness of the transformed structure reduces. Correspondingly, the velocity  $c$  of the acoustic wave increases. This gives rise to a correlation between hardness and the velocity of acoustic wave that can be used to determine the hardness nondestructively. The correlation in Fig. 35 was fitted with a second degree polynomial, excluding the data points for the non-hardened steels. The correlation coefficients and the standard deviations



**Fig. 35** Correlation of hardness HV20 with the velocity  $c$  of the surface acoustic wave at frequency  $f = 9$  MHz for laser-hardened steel 42CrMo4,  $T$  = annealing temperature (Schneider et al. 2012b, reprinted with permission of Elsevier, original Fig. 12)

are given. It should be mentioned here that a correlation of hardness  $H$  with Young's modulus  $E$  was already described for ta-C, shown in Fig. 15 (“[Super-Hard Tetrahedral Amorphous Carbon \(ta-C\)](#)”). Comparing Fig. 35 with Fig. 15 reveals that the  $H$ - $E$ -correlation is very different for steel and ta-C. For the ta-C films (Fig. 15), the correlation is positive,  $H \approx E/10$ . For the laser-hardened steel (Fig. 35), the correlation is negative, the higher the hardness  $H$ , the lower the acoustic wave velocity  $c$ . For the ta-C films, both hardness  $H$  and Young's modulus  $E$  vary within in a wide range,  $H$  from 12 to 50 GPa and  $E$  from 125 to 450 GPa (Fig. 15). In the case of the laser-hardened 42CrMo4, the hardness  $H$  also varies within a wide range, from 300 to 750 HV20, but the acoustic velocity  $c$  varies only from 2955 to 3005 m/s (Fig. 35). Considering Eq. 1,  $E \sim c^2 \rho$ , enables a variation of only 3.3% to be estimated for Young's modulus  $E$ . The conclusion can be drawn that there is no general rule for the correlation between hardness  $H$  and Young's modulus  $E$ . The  $H$ - $E$ -correlation depends on the material and the hardening mechanism. The martensitic hardening of the steel results in a fine micro- and sub-microstructure and a supersaturated solid solution of carbon in a body-centered tetragonal lattice (Kurdjumov 1960). These effects considerably reduces the plastic deformability, but do not influence significantly the metallic bonding. Martensitic hardening only expands the atomic lattice which reduces Young's modulus  $E$  only slightly. In contrast, ta-C films have an amorphous structure with a wide range of bonding conditions, specified by the ratio

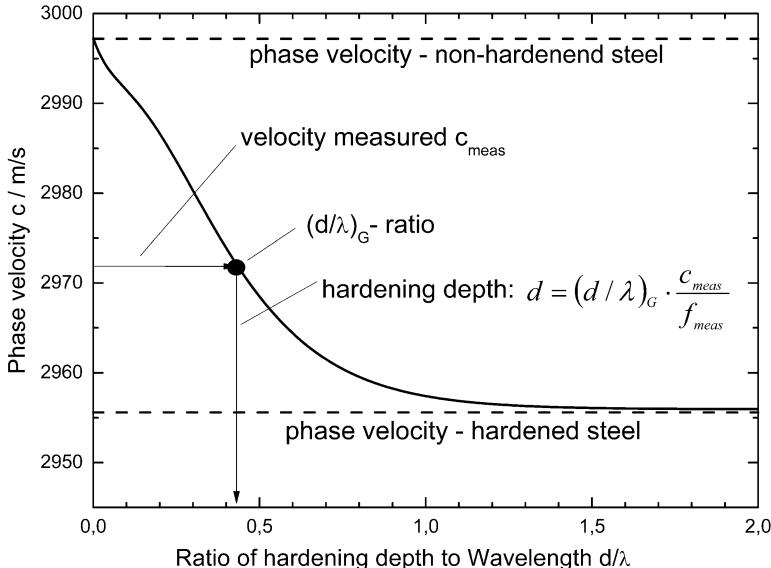


**Fig. 36** Hardening depth  $d_{LA}$  determined with the laser-acoustic method versus the hardening depth  $d_m$  obtained from metallography for the steel 42CrMo4, the lines for complete agreement and 10% deviation are given (Schneider et al. 2012b, reprinted with permission of Elsevier, original Fig. 16)

of  $sp^3/sp^2$ -bonds. The tetrahedral  $sp^3$  configuration of the carbon atoms is responsible for the extreme properties, such as high hardness  $H$  and high elastic modulus  $E$  (Robertson 2002). Therefore, there is a strong correlation between  $H$  and  $E$  for a large variation range. Similar correlations were found for hard coatings, such as carbides and nitrides (Jiang et al. 1991; Musil 2000; Musil et al. 2002; Rebholz et al. 2006).

Concerning the effect of the thickness of the hardened layer, the dispersion curves in Fig. 34 show that at low frequencies  $f \leq 4$  MHz the velocity  $c$  distinctly depends on the hardening depth. The higher the hardening depth, the lower is the velocity  $c$  of the surface acoustic wave. The effect of the hardening depth on the phase velocity  $c$  becomes the more pronounced, the lower the frequency  $f$ . The wave amplitude has an exponential depth profile. Therefore, the wave has the highest sensitivity to hardening depth for layers which have a depth in the range of half of the penetration depth of the wave. Figure 36 shows the correlation between the hardening depth  $d_{LA}$  determined with the surface acoustic wave method and the metallographic values  $d_m$  determined at the cross-sections of the laser-hardened steel 42CrMo4.

In order to determine the laser-acoustic hardening depth  $d_{LA}$ , a general curve was calculated for the laser-hardened steel 42CrMo4 which is shown Fig. 37 (Schneider



**Fig. 37** Generalized dispersion curve for the phase velocity of surface acoustic waves  $c$  versus the ratio of hardening depth to wavelength  $d/\lambda$  calculated for the steel 42CrMo4 (Schneider et al. 2012b, reprinted with permission of Elsevier, original Fig. 9)

et al. 2012b). The curve was calculated with a Young's modulus of  $E_0 = 211.2$  GPa for the non-hardened substrate and  $E_L = 205.4$  GPa for the hardened layer, using the theory in section “Surface Acoustic Waves.” For the substrate and the layer, the Poisson's ratio of  $\nu = 0.293$  and a density of  $\rho = 7.8$  g/cm<sup>3</sup> were taken from the data sheet. The Young's moduli of both materials were determined from the velocities measured at the non-hardened material and the laser-hardened surface, making use of Eq. 1.

The curve in Fig. 37 enables the hardness depth  $d$  to be determined in the following way. A dispersion curve  $c(f)$  as shown in Fig. 34 is measured at a hardened surface. For a specified frequency  $f_{\text{meas}}$ , the related velocity  $c_{\text{meas}}$  is taken from this curve  $c(f)$ . The velocity  $c_{\text{meas}}$  enables the related ratio  $(d/\lambda)_G$  to be found in the generalized dispersion curve (Fig. 37). Making use of Eq. 44, the hardening depth  $d$  can be calculated using the three parameters: frequency  $f_{\text{meas}}$ , velocity  $c_{\text{meas}}$ , and ratio  $(d/\lambda)_G$ .

$$d = (d/\lambda)_G \cdot \frac{c_{\text{meas}}}{f_{\text{meas}}} \quad (44)$$

The dispersion curve measured contains many data points  $c(f)$ . Therefore, the hardening depth  $d$  can be calculated for different  $f_{\text{meas}}$  in the range  $f \leq 4$  MHz, using Eq. 44. The mean value of  $d$  and the standard deviation are calculated.

The laser-acoustic measurements of the hardening depth in Fig. 36 were done with a 2.25 MHz sensor and a 10 MHz sensor. The 2.25 MHz sensor detects waves with deeper penetration depth than the 10 MHz sensor. Therefore, its depth range is larger. The method is well suited for testing hardened layers in the range of up to about one millimeter. To penetrate deeper into the material, the frequency must be decreased.

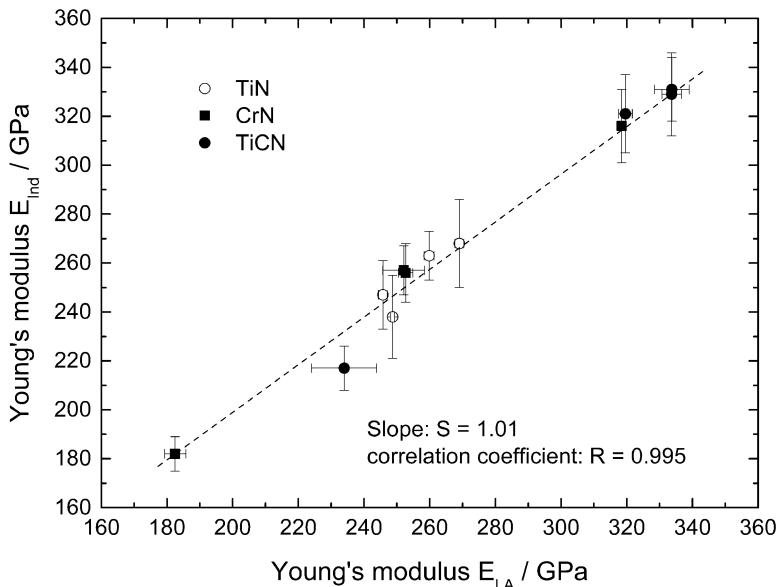
---

## Comparing the Results with Other Methods

Several methods are used for measuring the elastic moduli of thin films and coatings. They can be divided into quasi-static methods, such as membrane-deflection (Maier-Schneider et al. 1995), instrumented micro- and nano-indentation (Oliver and Pharr 1992), and dynamic methods, such as resonance ultrasound spectroscopy, Brillion scattering (Zinin et al. 2002), acoustic microscopy (Zinin et al. 1997), and laser-acoustics (Hess 2002, 2009). The instrumented deep-sensing indentation is the most commonly used technique. Therefore, it is important to compare the results of this quasi-static indentation test with results obtained with surface acoustic waves.

Figure 38 shows the values  $E_{Ind}$  measured with the micro-indenter Fischerscope H100VP-B versus the values  $E_{LA}$  of Young's modulus measured with the laser-acoustic technique, which were determined for different hard coatings, TiN, TiCN, and CrN (Schneider et al. 1998a). The films were deposited by magnetron sputtering on steel. The film thickness was in the range of 0.8–2.3  $\mu\text{m}$ , which allows the conditions for correct indentation tests to be met. The indentation test was performed with a load of 20 mN to make sure that the effect of the substrate is excluded. The tests were repeated five times with both test methods. In Fig. 38, the mean values and the standard deviations are plotted. This comparison test shows a very good agreement of the results of these very different test methods which is proven by a slope of the linear regression of  $S = 1.01$  and a correlation coefficient of  $R = 0.995$ .

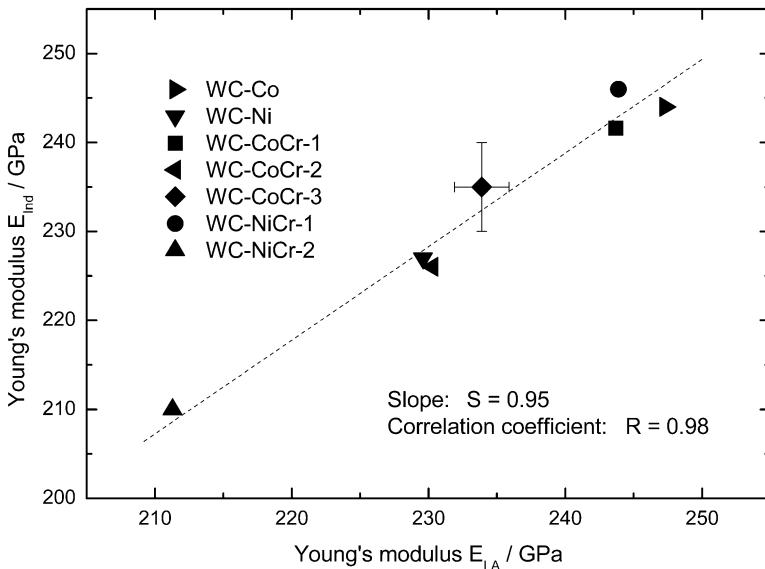
Figure 39 shows the comparison test of indentation and surface acoustic waves which was performed at tungsten-carbide-based hard-metal coatings which were produced by high-velocity oxy-fuel (HVOF-) spraying (Berger et al. 2007). Seven different spray powders were used to spray the coatings on mild steel substrates. Metal powders of Ni, Co, and Cr were added to the WC feed stock which serve as binder for the WC particles. The thickness of the coatings varied from  $d = 200 \mu\text{m}$  to  $350 \mu\text{m}$ . The dispersion of the surface acoustic wave was measured with the device described in section “[Experimental Setup and Measuring Procedure](#)” on the as-sprayed surface of the coatings with a roughness from  $R_a = 3.9$  to  $6.0 \mu\text{m}$ . The micro-indentation tests were performed at metallographically prepared cross-section of the coatings, using a Shimadzu DUH-202 instrument (load: 0.5 N, Vickers indents). The hardness and Young's modulus were determined from the force-



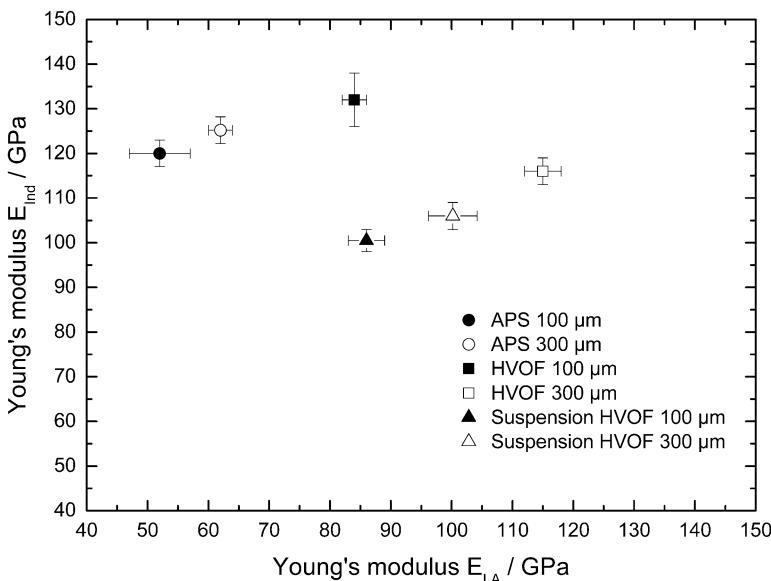
**Fig. 38** Comparison of Young's modulus  $E_{LA}$  determined by the laser-acoustic technique and the micro-indentation  $E_{Ind}$  for films of TiN, TiCN, and CrN deposited on steel (Schneider et al. 1998, reprinted with permission of Elsevier, original Fig. 9)

displacement curve yielded by the instrument (Oliver and Pharr 1992). Figure 39 shows also for this application a good agreement between the results of both methods. The linear regression results in a slope of  $S = 0.95$  and a correlation coefficient of  $R = 0.98$ . The metallographic investigation proved that compact WC-based coatings were deposited which had a globular porosity between 0.7% and 6.6% (Berger et al. 2007).

Figure 40 shows the comparison of the surface acoustic wave method and the indentation test for another example of thermal-sprayed coatings, the  $\text{Al}_2\text{O}_3$  coatings described in section “[Thermal-Sprayed Coatings](#).” For these coatings, the surface wave test was also performed on the as-sprayed surface and the micro-indentation tests were performed on metallographically prepared cross-sections with the Shimadzu DUH-202 tester (load: 0.5 N, Vickers indents) (Berger et al. 2011). Contrary to the results for the tungsten-carbide-based coatings shown in Fig. 39, the correlation between the results of the two techniques is poor. The values of Young's modulus  $E_{LA}$  measured with the laser-acoustic technique vary from 54 to 116 GPa, whereas the values  $E_{Ind}$  measured with the indentation tester are in the range from 100 to 132 GPa. The best agreement was observed for the Suspension HVOF coatings and HVOF coatings with a thickness  $d = 300 \mu\text{m}$ . For the APS coatings, the value obtained with the indenter is almost twice as high as that yielded from the surface acoustic wave method. The differences require an explanation. The surface acoustic wave method is a more global test, evaluating a larger volume of the material compared



**Fig. 39** Comparison of Young's modulus  $E_{LA}$  determined by the laser-acoustic technique and the micro-indentation  $E_{Ind}$  for WC-based thermal-sprayed coatings,  $E_{LA}$  was measured at the surface and  $E_{Ind}$  was measured at the metallographic cross-section of the coatings (Berger et al. 2007)



**Fig. 40** Comparison of Young's modulus  $E_{LA}$  determined by the laser-acoustic technique and the micro-indentation  $E_{Ind}$  for  $\text{Al}_2\text{O}_3$  coatings with different thickness  $d$  sprayed with APS, HVOF and Suspension HVOF (Berger 2011)

to the micro-indenter. The highest frequency  $f = 6$  MHz in the dispersion curves in Fig. 21 has the shortest wavelength,  $\lambda = 500$   $\mu\text{m}$ . Therefore, the test dimension is large compared to the defective microstructure shown in the metallographic picture in Fig. 25. Young's modulus derived from the surface acoustic waves for these coarsely structured coatings is an effective modulus in the understanding of theory of heterogeneous materials introduced in section “[Thermal-Sprayed Coatings](#).” The value represents all microstructural components and defects within a comparably large volume. In contrast, the indentation test measures Young's modulus  $E_{Ind}$  within a comparably small and local area at the metallographic cross-section of the coatings. Any single tests showing a load-displacement curve that did not meet the requirements of the theoretical model were excluded from the statistics. In this way, material regions with pores and micro-cracks did not contribute to the result in its entirety. Therefore, the indenter is expected to represent more the  $\text{Al}_2\text{O}_3$  matrix and to a lesser extend the effective Young's modulus for materials with many pores and micro-cracks.

---

## Summary

Surface acoustic waves are a nondestructive test method which can successfully be used for films and coatings in the range from a few nanometers up to millimeter by adapting the frequency range.

An experimental setup was described that uses laser pulses with the duration of 3 ns to generate wideband acoustic pulses which are received by a piezoelectric sensors with a frequency range up to 250 MHz. For rough surfaces and coarse-grained materials, a 10 MHz sensor is used. It can measure the phase velocity as a function of frequency  $c(f)$  (dispersion curve) with an uncertainty of  $\Delta c/c \leq \pm 2 \times 10^{-4}$ .

Two theoretical approaches were described which can be used for the determination of Young's modulus  $E$ , density  $\rho$ , and thickness  $d$  of the coating from the measured dispersion curve by a curve fitting procedure, the global matrix and the transfer matrix model.

It was demonstrated how the form of the dispersion curve depends on the properties of the film and the substrate material. The substrate material determines the zero point of the dispersion curve. The film determines the shape of the dispersion curve. For a very thin film, the dispersion curve, which can be measured by the experimental setup, is linear. In this case, only one film parameter, e.g., Young's modulus  $E$ , can be calculated. Film thickness  $d$  and density  $\rho$  must be measured with alternative methods. Increasing the film thickness causes a curvature of the dispersion curve, which allows more parameters to be calculated. It depends on the combination of substrate and film material, the film thickness, and the bandwidth of the measurement how many film parameters can be determined. Surface roughness, grain scattering, and internal friction reduce the bandwidth.

Five examples of application were presented:

1. Young's modulus of super-hard ta-C films with a film thickness down to few nanometers was measured. It correlates with the content of  $sp^3$  bonds and therefore also with the hardness of this material.
2. The example of the titanium films demonstrate that porosity distinctly reduces the Young's modulus compared to the value known for the bulk material. Therefore, the test method was helpful for optimizing the technology to deposit dense films.
3. For thermal-sprayed coatings, Young's modulus can be measured on as-sprayed surfaces which have a roughness up to more than  $R_a = 6.5 \mu\text{m}$ . The frequency range of the test must be adapted to  $f \leq 10 \text{ MHz}$ . Young's moduli measured for sprayed  $\text{Al}_2\text{O}_3$  coatings were much lower ( $E < 150 \text{ GPa}$ ) than the value known for the bulk material ( $E = 350 \text{ GPa}$ ), due to the effect of micro-defects.
4. Cutting a semiconductor wafers from ingots produces a zone of high defect density, the subsurface damage. It was demonstrated that surface acoustic waves can be used for the evaluation of the depth of the subsurface damage. For this application, the method has to be calibrated with reference samples.
5. Martensitic hardening of a surface layer in steel is accompanied by a reduction of the velocity of the surface acoustic wave. Annealing experiments showed that there is a correlation between decreasing surface hardness and increasing velocity, which can be used for nondestructive evaluation of the hardening state. For the laser-hardened steel 42CrMo4, it was demonstrated that the hardening depth can be determined up to the depth of 1 mm.

Results for the Young's modulus obtained with the surface acoustic wave technique were compared with the instrumented indentation test. Both methods showed good agreement for dense and compact films and coatings. For defective materials as thermal-sprayed ceramics, the surface wave method measured considerably lower values than indentation test.

---

## References

- Akhmanov SA, Gusev VE (1992) Laser excitation of ultrashort acoustic pulses: new possibilities in solid-state spectroscopy, diagnostics of fast processes, and nonlinear acoustics. Sov Phys Usp 35:153–191
- Arnold W, Betz B, Hoffmann B (1985) Efficient generation of surface acoustic waves by thermoelasticity. Appl Phys Lett 47:672–674. <https://doi.org/10.1063/1.96054>
- Asmani M, Kermel C, Leriche A, Ourak MJ (2001) Influence of porosity on Young's modulus and Poisson's ratio in alumina. Eur Ceram Soc 21:1081–1086
- Aussel JD, Monchalin J (1989) Precision laser-ultrasonic velocity measurement and elastic constant determination. Ultrasonics 27:165–177. [https://doi.org/10.1016/0041-624X\(89\)90059-0](https://doi.org/10.1016/0041-624X(89)90059-0)
- Bennis A, Lomonosov M, Shen ZH, Hess P (2006) Laser-based measurement of elastic and mechanical properties of layered polycrystalline silicon structures with projection masks. Appl Phys Lett 88:101915-1–101915-3. <https://doi.org/10.1063/1.2181187>
- Berger LM, Schneider D, Großer T (2007) Non-destructive testing of coatings by surface acoustic waves. In: Marple BR, Hyland MM, Lau YC, Li CJ, Lima RS, Montavon G (eds) Thermal spray 2007: global coating solutions. ASM International, Materials Park, pp 916–921

- Berger LM (2011) Entwicklung einer zerstörungsfreuen Prüfmethode zur Messung mechanischer Kennwerte und der Porosität an thermisch gespritzten Schichten. Final report of the IGF Project 16.029 BR / DVS No. 02.056, promoted by the German Ministry of Economic Affairs and Technology (BMWi) via AiF within the framework of the program for the promotion of joint industrial research and development, Fraunhofer Institute for Material and Beam Technology (IWS) Dresden
- Berger LM, Schneider D, Barbosa M, Puschmann R (2012) Laser acoustic surface waves for the non-destructive characterization of thermally sprayed coatings. *Therm Spray Bull* 64(1):56–64
- Bescond C, Kruger SE, Le'veque D, Lima RS, Marple BR (2007) In-situ simultaneous measurement of thickness, elastic moduli and density of thermal sprayed WC-Co coatings by Laser-Ultrasonics. *J Therm Spray Technol* 16:238–244. Thermal spray coatings
- Carvalho S, Vaz F, Rebouta L, Schneider D, Cavaleiro A, Alves E (2001) Elastic properties of (Ti,Al,Si)N nanocomposite films. *Surf Coat Technol* 142–144:110–116
- Coufal H, Grygier R, Hess P, Neubrand A (1992) A broadband detection of laser-excited surface acoustic waves by a novel transducer employing ferroelectric polymers. *J Acoust Soc Am* 92:2980–2983
- Dobmann G, Kern R, Altpeter I, Theiner W (1988) Quantitative hardening-depth-measurements up to 4mm by means if micro-magnetic microstructure multi-parameter analysis (3MA). In: Thomson DO, Chimenti DE (eds) Review of progress in quantitative nondestructive evaluation, vol 7b. Springer, Boston, pp 1471–1475
- Eshelby JD (1957) The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proc. Royal Soc A* 241:376–396. <https://doi.org/10.1098/rspa.1957.0133>
- Farnell GW (1970) Properties of elastic surface waves. In: Mason WP, Thurston PM (eds) Physical acoustics, vol VI. Academic, New York/London, pp 109–166
- Farnell GW, Adler EL (1972) Elastic wave propagation in thin layers. In: Mason WP, Thurston RN (eds) Physical acoustics, vol IX. Academic, New York/London, pp 35–127. <https://doi.org/10.1016/B978-0-12-395670-5.50007-6>
- Flannery CM, Murray C, Streiter I, Schulz SE (2001) Characterization of thin-film aerogel porosity and stiffness with laser-generated surface acoustic waves. *Thin Solid Films* 388:1–4
- Grünwald E, Nuster R, Treml R, Kiener D, Paltauf G, Brunner R (2015) Young's Modulus and Poisson's ratio characterization of tungsten thin films via laser ultrasound. *nanoFIS 2014 Mater Today* 2:4289–4294
- Hashin ZJ (1962) The elastic moduli of heterogeneous materials. *Appl Mech* 29:143–150
- Haskell NA (1953) The dispersion of surface waves on multilayered media. *Bull Seismol Soc Am* 43:17–34
- Hess P (2002) Surface acoustic waves in materials science. *Phys Today* 55:42:47
- Hess P (2009) Determination of linear and nonlinear mechanical properties of diamond by laser-based surface acoustic waves. *Diamond Relat Mater* 18:186–190. <https://doi.org/10.1016/j.diamond.2008.10.005>
- Hill R (1965) Continuum micro-mechanics of elastoplastic polycrystals. *J Mech Phys Solids* 13:89–101
- Jiang X, Wang M, Schmidt K, Dunlop E, Haupt J, Gissler W (1991) Elastic constants and hardness of ion beam sputtered TiNx films measured by Brillouin scattering and depth sensing indentation. *J Appl Phys* 69:3053–3057. <https://doi.org/10.1063/1.348963>
- Karabutov AA (1985) Laser excitation of surface acoustic waves: a new direction in opto-acoustic spectroscopy of a solid. *Sov Phys Usp* 28:1042–1051
- Knopoff L (1964) A matrix method for elastic wave problems. *Bull Seismol Soc Am* 54:431–438
- Kolomenskii AA, Szabadi M, Hess P (1995) Laser diagnostics of C60 and C70 films by broadband surface acoustic wave spectroscopy. *Appl Surf Sci* 86:591–596
- Kreher W, Janssen R (1992) On microstructural residual stresses in particle reinforced ceramics. *J Eur Ceram Soc* 10:167–173
- Kreher W, Pompe W (eds) (1989) Internal stresses in heterogeneous solids. Akademie-Verlag, Berlin
- Kröner E (1961) Zur plastischen Verformung des Vielkristalls. *Acta Metall* 9:155–161

- Kurdjumov GV (1960) Phenomena occurring in the quenching and tempering of steels. *J Iron Steel Inst* 195:26–48
- Kuschnerit R, Fath H, Kolomenskii AA, Szabadi M, Hess P (1995) Mechanical and elastic properties of amorphous hydrogenated silicon films studied by broadband surface acoustic wave spectroscopy. *Appl Phys A* 61:269–276. <https://doi.org/10.1007/BF01538192>
- Landau LD, Lifschitz EM (1975) Lehrbuch der theoretischen Physik. vol. 7. Elastizitätstheorie. Akademie Verlag, Berlin
- Lee RE, White RM (1968) Excitation of surface elastic waves by transient surface heating. *Appl Phys Lett* 12:12–14. <https://doi.org/10.1063/1.1651832>
- Leonhardt M, Schneider D, Kaspar J, Schenk S (2004) Characterizing the porosity in thin titanium films by laser-acoustics. *Surf Coat Technol* 185:292–302. <https://doi.org/10.1016/j.surfcoat.2004.01.020>
- Lima RS, Kruger SE, Lamouche G, Marple BR (2005) Elastic Modulus measurements via laser-ultrasonic and Knoop indentation. *J Therm Spray Technol* 14:52–60. <https://doi.org/10.1361/10599630522701>
- Lomonosov AM, Mayer AP, Hess P (2001) Laser-based surface acoustic waves in material science. In: Levy M, Bass HE, Stern R (eds) Modern acoustical techniques for the measurement of mechanical properties, vol 39. Academic, San Diego, pp 65–134
- Lowe MJS (1995) Matrix techniques for modeling ultrasonic waves in multilayered media. *IEEE Trans Ultrason Ferroelectr Freq Control* 42:525–542. <https://doi.org/10.1109/58.393096>
- Lyamshev LM (1981) Optoacoustic sources of sound. *Sov Phys Usp* 24:977–995
- Maier-Schneider D, Ersay A, Maibach J, Schneider D, Obermeier E (1995) Influence of annealing on the elastic properties of LPCVD silicon nitride and LPCVD polysilicon. *Sens Mater* 7:121–129
- Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math* 11(2):431–441
- Maznev AA, Mazurenko A, Li Z, Gostein M (2003) Laser-based surface acoustic wave spectrometer for industrial applications. *Rev Sci Instrum* 74:667–669. <https://doi.org/10.1063/1.1512680>
- Mittal KL (ed) (1976) Adhesion measurement of thin films, thick films and bulk coatings, ASTM Symposium Philadelphia, ASTM special technical publication, vol 640. ASTM, Philadelphia
- Monchal JP (1985) Optical detection of ultrasound at a distance using a confocal Fabry–Perot interferometer. *Appl Phys Lett* 47:14–16. <https://doi.org/10.1063/1.96411>
- Musil J (2000) Hard and superhard nanocomposite coatings. *Surf Coat Technol* 125:322–330. [https://doi.org/10.1016/S0257-8972\(99\)00586-1](https://doi.org/10.1016/S0257-8972(99)00586-1)
- Musil J, Kunc F, Zeman H, Polakova H (2002) Relationships between hardness, Young's modulus and elastic recovery in hard nanocomposite coatings. *Surf Coat Technol* 154:304–313. [https://doi.org/10.1016/S0257-8972\(01\)01714-5](https://doi.org/10.1016/S0257-8972(01)01714-5)
- Neubrand A, Hess P (1992) Laser generation and detection of surface acoustic waves: elastic properties of surface layers. *J Appl Phys* 71:227–238
- O'Connor DJ, Sexton BA, Smart R, St C (eds) (1992) Surface analysis methods in materials science. Springer, Heidelberg
- Oliver WC, Pharr GM (1992) An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. *J Mater Res* 7:1564–1583
- Ollendorf H, Schneider D (1999) A comparative study of adhesion test methods for hard coatings. *Surf Coat Technol* 113:86–102
- Paehter D, Schneider D, Herben M (2007) Nondestructive characterization of sub-surface damage in rotational ground silicon wafers by laser acoustics. *Microelectron Eng* 84:340–354. <https://doi.org/10.1016/j.mee.2006.11.001>
- Rayleigh L (1885) On waves propagating along the plane surface of an elastic solid. *Proc Lond Math Soc* 17:4
- Rebholz C, Leyland A, Matthews A, Charitidis C, Logothetidis S, Schneider D (2006) Correlation of elastic modulus, hardness and density for sputtered TiAlBN thin films. *Thin Solid Films* 514:81–86
- Retzko I, Unger W (2003) Analysis of carbon materials by X-ray photoelectron spectroscopy and X-ray absorption spectroscopy. *Adv Eng Mater* 5:519–522. <https://doi.org/10.1002/adem.200320138>

- Robertson J (2002) Diamond-like amorphous carbon. *Mater Sci Eng R* 37:129–281
- Rogers JA, Yang Y, Nelson KA (1994) Elastic Modulus and In-Plane Thermal Diffusivity Measurements in Thin Polyimide Films Using Symmetry-Selective Real-Time Impulsive Stimulated Thermal Scattering. *Appl Phys A* 58:523–534
- Royer D (2001) Mixed matrix formulation for the analysis of laser-generated acoustic waves by thermoelastic line sources. *Ultrasonics* 39:345–354
- Royer D, Chenu C (2000) Experimental and theoretical waveforms of Rayleigh waves generated by a thermoelastic laser line source. *Ultrasonics* 38:891–895
- Ruiz AM, Nagy PB (2002) Diffraction correction for precision surface acoustic wave. *J Acoust Soc Am* 112(3):835–842. <https://doi.org/10.1121/1.1497368>
- Sachse W, Pao YH (1978) On the determination of phase and group velocities of dispersive waves in solids. *J Appl Phys* 49:4320–4327. <https://doi.org/10.1063/1.325484>
- Schneider D (2013) Laser acoustic testing machine for the surface analysis of silicon blocks and solar wafers. *Annu Rep Fraunhofer IWS Dresden*, pp 52–53
- Schneider D, Franke K (1990) Anwendung interdigitaler Oberflächenwellenwandler für die zerstörungsfreie Werkstoffprüfung. *Feingerätetechnik* 39:117–120
- Schneider D, Schwarz T (1997) A photoacoustic method for characterizing thin films. *Surf Coat Technol* 91:136–146
- Schneider D, Herrmann K, Brenner B, Schläfer D, Winderlich B (1986) Investigation of the influence of grinding on regions near the surface by ultrasonic surface waves. *Cryst Res Technology* 21:897–905
- Schneider D, Schwarz T, Schultrich B (1992) Determination of elastic modulus and thickness of surface layers by ultrasonic surface waves. *Thin Solid Films* 219:92–102
- Schneider D, Schwarz T, Buchkremer HP, Stöver D (1993) Non-destructive characterization of plasma-sprayed ZrO<sub>2</sub> coatings by ultrasonic surface waves. *Thin Solid Films* 224:177–183
- Schneider D, Schwarz T, Scheibe HJ, Panzner M (1997) Non-destructive evaluation of diamond and diamond-like carbon films by laser induced surface acoustic waves. *Thin Solid Films* 295:107–116
- Schneider D, Schultrich B, Scheibe HJ, Ziegele H, Grieppentrog M (1998a) A laser-acoustic method for testing and classifying hard surface layers. *Thin Solid Films* 332:157–163
- Schneider D, Meyer CF, Mai H, Schöneich B, Ziegele H, Scheibe HJ, Lifshitz Y (1998b) Nondestructive evaluation of diamond and diamond-like carbon films by laser induced surface acoustic waves. *Diam Relat Mater* 7:973–980
- Schneider D, Hammer R, Jurisch M (1999) Non-destructive testing of damage layers in GaAs wafers by surface acoustic waves. *Semicond Sci Technol* 14:93–98. <https://doi.org/10.1088/0268-1242/14/1/015>
- Schneider D, Siemroth P, Schuelke T, Berthold J, Schultrich B, Schneider HH, Ohr R, Peterit B, Hillgers H (2002a) Quality control of ultra-thin and super-hard coatings by laser-acoustics. *Surf Coat Technol* 153:252–260. [https://doi.org/10.1016/S0257-8972\(01\)01664-4](https://doi.org/10.1016/S0257-8972(01)01664-4)
- Schneider D, Stiehl E, Hammer R, Franke A, Riegert R, Schuelke T (2002b) Nondestructive testing of damage layers in semiconductor materials by surface acoustic waves. *Proc SPIE* 4692:195–203. <https://doi.org/10.1117/12.475660>
- Schneider D, Frühauf S, Schulz SE, Gessner T (2005) The current limits of the laser-acoustic test method to characterize low-k films. *Microelectron Eng* 82:393–398. <https://doi.org/10.1016/j.mee.2005.07.073>
- Schneider D (2012) Using laser induced surface acoustic waves to characterize thin films and material surfaces. *IEEE International Ultrasonics Symposium, IUS* pp. 269–272. <https://doi.org/10.1109/ULTSYM.2012.0066>
- Schneider D, Leson A, Berger LM (2012a) Laserakustik für Schicht- und Oberflächenprüfung. *Vakuum in Forschung und Praxis* 24(4):17–23. <https://doi.org/10.1002/vipr.201200499>
- Schneider D, Hofmann R, Schwarz T, Grosser T, Hensel E (2012b) Evaluating surface hardened steels by laser-acoustics. *Surf Coat Technol* 206:2079–2088. <https://doi.org/10.1016/j.surfcoat.2011.09.017>

- Schuelke T, Anders A, Siemroth P (1997) Macroparticle filtering of high-current vacuum arc plasmas. *IEEE Trans Plasma Sci* 25:660–664. <https://doi.org/10.1109/27.640681>
- Schuelke T, Witke T, Scheibe HJ, Siemroth P, Schultrich B, Zimmer O, Vetter J (1999) Comparison of DC and AC arc thin film deposition techniques. *Surf Coat Technol* 120–121:226–232
- Schultrich B, Scheibe HJ, Grandremy G, Drescher D, Schneider D (1996) Elastic modulus as a measure of the diamond likeness and hardness of amorphous carbon films. *Diam Relat Mater* 5:914–918
- Schulz H (2005) Amorphe Kohlenstoffschichten hergestellt mittels Laser-Arc Verfahren unter besonderer Berücksichtigung der Oberflächentopographie zur Herstellung superhydrophober Oberflächen. PhD Thesis, TU Dresden, Fakultät Maschinenwesen
- Shan Q, Jawad SM, and Dewhurst RJ (1993) An automatic stabilization system for a confocal Fabry-Perot interferometer used in the detection of laser-generated ultrasound. *Ultrasonics* 31:105–115
- Silva SR, Xu S, Tay BK, Tan HS, Scheibe HJ, Chhowalla M, Milne WI (1996) The structure of tetrahedral amorphous carbon thin films. *Thin Solid Films* 290:317–322
- Singer F, Kufner M (2017) Model based laser-ultrasound determination of hardness gradients of gascarburized steel. *NDT&E Int* 88:24–32
- Szabo TL, Slobodnik AL (1973) The effect of diffraction on the design of acoustic surface wave devices. *IEEE Trans Sonics Ultrason* 20:240–251
- Thomson WT (1950) Transmission of elastic waves through a stratified solid medium. *J Appl Phys* 21:89–93. <https://doi.org/10.1063/1.1699629>
- White RM (1963) Generation of elastic waves by transient surface heating. *J Appl Phys* 34:3559–3567. <https://doi.org/10.1063/1.1729258>
- White RM (1970) Surface elastic waves. *Proc IEEE* 58:1238–1276
- Whitman RL, Korpel A (1969) Probing of acoustic surface perturbations by coherent light. *Appl Opt* 8:1567–1576
- Wienss A, Persch-Schuy G, Vogelgesang M, Hartmann U (1999) Scratching resistance of diamond-like carbon coatings in the subnanometer regime. *Appl Phys Lett* 75:1077:1079
- Willems H (1991) Nondestructive determination of hardening depth in induction hardened components by ultrasonic backscattering. In: Thompson DO, Chimenti DE (eds) Review of progress in quantitative non-destructive evaluation, vol 10B. Plenum Press, New York, pp 1707–1713
- Xiao X, You X (2006) Numerical study on surface acoustic wave method for determining Young's modulus of low-k films involved in multi-layered structures. *Appl Surf Sci* 253:2958–2963
- Xiao X, Qi H, Tao Y, Kikkawal T (2016) Study on the interfacial adhesion property of low-k thin film by the surface acoustic waves with cohesive zone model. *Appl Surf Sci* 388:448–454
- Zinin P, Lefevre O, Briggs GAD, Zeller D, Cawley P, Kinloch AJ (1997) Anomalous behaviour of leaky surface waves for stiffening layer near cutoff. *J Appl Phys* 82:1031–1035
- Zinin P, Manghnani MH, Zhang X, Feldermann H, Ronning C, Hofstass H (2002) Surface Brillouin scattering of cubic boron nitride films. *J Appl Phys* 91:4196–4204

## Reprinted from the Following Publications with Permission from Elsevier

- Leonhardt M, Schneider D, Kaspar J, Schenk S (2004) Characterizing the porosity in thin titanium films by laser-acoustics. *Surf Coat Technol* 185:292–302. Elsevier Reuse License Number 4186960496489
- Paehler D, Schneider D, Herben M (2007) Nondestructive characterization of sub-surface damage in rotational ground silicon wafers by laser acoustics. *Microelectron Eng* 84:340–354. <https://doi.org/10.1016/j.mee.2006.11.001>. Elsevier Reuse License Number 4187020428836

- 
- Schneider D, Schwarz T (1997) A photoacoustic method for characterizing thin films. *Surf Coat Technol* 91:136–146. Elsevier Reuse License Number 4160710855584
- Schneider D, Schwarz T, Scheibe HJ, Panzner M (1997) Non-destructive evaluation of diamond and diamond-like carbon films by laser induced surface acoustic waves. *Thin Solid Films* 295:107–116. Elsevier Reuse License Number 4164621126820
- Schneider D, Schultrich B, Scheibe HJ, Ziegele H, Griepong M (1998) A laser-acoustic method for testing and classifying hard surface layers. *Thin Solid Films* 332:157–163. Elsevier Reuse License Number 4187021304650
- Schneider D, Siemroth P, Schuelke T, Berthold J, Schultrich B, Schneider HH, Ohr R, Petereit B, Hillgers H (2002) Quality control of ultra-thin and super-hard coatings by laser-acoustics. *Surf Coat Technol* 153:252–260. [https://doi.org/10.1016/S0257-8972\(01\)01664-4](https://doi.org/10.1016/S0257-8972(01)01664-4). Elsevier Reuse License Number 4186911186263
- Schneider D, Frühauf S, Schulz SE, Gessner T (2005) The current limits of the laser-acoustic test method to characterize low-k films. *Microelectron Eng* 82:393–398. <https://doi.org/10.1016/j.mee.2005.07.073>. Elsevier Reuse License Number 4186910426737
- Schneider D, Hofmann R, Schwarz T, Grosser T, Hensel E (2012b) Evaluating surface hardened steels by laser-acoustics. *Surf Coat Technol* 206:2079–2088. <https://doi.org/10.1016/j.surfcoat.2011.09.017>. Elsevier Reuse License Number 4187020878545



# The Acousto-elastic Effect and Its Use in NDE

7

Hans-Rüdiger Herzer, Michael Mathias Becker, and Eckhardt Schneider

## Contents

Introduction .....	236
Basics .....	236
Applications .....	238
Evaluation of a One-Axial Stress State .....	238
Evaluation of a Two-Axial Stress State .....	241
Evaluation of a Three-Axial Stress State .....	246
Fields for Further Development of Ultrasonic Stress Evaluation .....	247
Summary .....	248
Evaluation of Clamp Load of Bolts .....	248
Evaluation of Stress States of Railroad Wheels .....	249
Other Applications .....	249
References .....	250

## Abstract

The acousto-elastic effect describes the influence of an elastic strain state on the elastic properties of a metallic material and hence on the velocities of elastic waves propagating in the strained component. The effect is used to evaluate stress states by ultrasonic methods. This chapter summarizes the fundamental concepts, measuring techniques, and evaluation procedures of nondestructive ultrasonic methods to evaluate one, two, and three axial stress states of components. The state of the art is detailed by discussing results of applications in, e.g., bolts, sheets, gear parts, and turbine rotors. The applicability and limitations of such techniques will be discussed in detail.

---

H.-R. Herzer (✉) · M. M. Becker · E. Schneider

Fraunhofer Institute for Nondestructive Testing IZFP, Saarbrücken, Germany

e-mail: [Hans-Ruediger.Herzer@izfp.fraunhofer.de](mailto:Hans-Ruediger.Herzer@izfp.fraunhofer.de); [Michael.Becker@izfp.fraunhofer.de](mailto:Michael.Becker@izfp.fraunhofer.de);  
[EckhardtSchneider@gmx.de](mailto:EckhardtSchneider@gmx.de)

## Introduction

In the period between 1970 and 1990, a large number of publications described the physical background and different applications of ultrasonic testing (see ► [Chap. 1, "Physical Basis for Ultrasonic Acoustics"](#)). Thompson et al. and Schneider (Thompson et al. 1996; Schneider 1997, 2000) described the background and summarized the state which, besides the development of robust systems, has not significantly changed up to now. In recent years, a few individual researchers are stepping into that area of application-related research and are reinvestigating more or less known subjects. That said, only very few groups have developed systems which are applied in the industrial environment. It is the intention of this contribution to inform on the states of the art with regard to the application on components and to support the user's decision to apply the appropriate technique.

Complementary to the established techniques to evaluate stress states, ultrasonic techniques permit the evaluation of stresses in surface layers and in the bulk of components, respectively. One advantage of ultrasonic techniques is the possibility to attain a fast evaluation of stress states, enabling a continuous analysis along measuring traces to get information on the stress distribution and to localize the stress inhomogeneity. Data rates are typically between about 1 and 50 Hz depending on the application and the ultrasonic setup used. Resolution and accuracy of the nondestructive ultrasonic techniques are of about the same order of magnitude as of the partly destructive drilling hole and ring core techniques. However, a disadvantage is that the evaluated stress result is a mean value of the particular stress component acting along the ultrasonic path length. The evaluation of the stress gradient along the path length is usually not possible.

The quantitative evaluation of stress states using ultrasonic techniques assumes the knowledge of the material-dependent acousto-elastic constants, which are usually evaluated by tensile testing. Assuming the acousto-elastic constants are known, the choice and the adaptation of the sensor has to be considered depending on the stress state to be evaluated, the geometry of the component or the accessibility to apply the sensors to the component.

To overcome restrictions based on geometry, inhomogeneity, accessibility, and ultrasonic hardware, methods are required which include information about the component's mechanical, chemical, and microstructural properties as well as production steps.

---

## Basics

The velocities of ultrasonic waves propagating in a component are influenced by the strain state of the component. This so-called acousto-elastic effect is the major principle of all ultrasonic techniques to evaluate strain or stress states. Using the fundamentals of the finite deformation of an elastic solid, described by F. D. Murnaghan (1951) and the description of the Second-Order Elastic Deformation of

Solids by D. S. Hughes and J. L. Kelly (1953), the relative changes of the propagation velocities of ultrasonic waves can be expressed as a function of the material-dependent acousto-elastic constants and the three components of the strain state. Assuming that the principal directions of strain go along with the principal directions of the stress state, the influence of stress on the relative change of the ultrasonic velocities can be described as (Schneider 1997, 2000).

$$\frac{v_{ii} - v_L}{v_L} = \frac{A}{C} \cdot \sigma_i + \frac{B}{C} \cdot (\sigma_j + \sigma_k) \quad (1)$$

$$\frac{v_{ij} - v_T}{v_T} = \frac{D}{K} \cdot \sigma_i + \frac{H}{K} \cdot \sigma_j + \frac{F}{K} \cdot \sigma_k \quad (2)$$

$$\frac{v_{ik} - v_T}{v_T} = \frac{D}{K} \cdot \sigma_i + \frac{H}{K} \cdot \sigma_k + \frac{F}{K} \cdot \sigma_j \quad (3)$$

$\sigma_i$ ,  $\sigma_j$  and  $\sigma_k$  are the three components of a normalized stress tensor.  $i, j, k$  are the axes of a Cartesian coordinate system.  $v$  is the ultrasonic velocity, the first index represents the direction of propagation, the second the direction of vibration.  $v_L$  and  $v_T$  stand for the longitudinal wave velocity and shear wave velocity in the stress free state, respectively.  $A, B, C, D, H, F$  and  $K$  are combinations of second (Young's- and shear modulus) and third order elastic (Murnaghan) constants (Schneider 2000). They are weighting the influence of the principal stress component on the change of the ultrasonic velocities. The material-dependent third-order elastic constants are experimentally evaluated using a sample of the material of interest as described by D. M. Egle and D. E. Bray (1976). Second and third order constants of some steel grades and Al alloys are given in several publications (Thompson et al. 1996; Schneider 1997, 2000; Egle and Bray 1976). Using the constants and the evaluation equations as given in Schneider (2000), the abovementioned weighting factors A through K are calculated. Hence, the influence of any stress state on the velocities of ultrasonic waves propagating and vibrating in different directions can be calculated. It is important to note that the Eqs. (1, 2, and 3) can only be applied if the directions of ultrasonic propagation and vibration are parallel to the principal strain and stress directions.

It has been found that the elastic constants of different ferritic steel grades are very similar and the influence of rolling and hardening, as well as the influence of the welding process on the elastic properties of the steel samples, is not as significant as it was initially assumed. Hence, using averaged values, the ultrasonic stress analysis yields a rough estimation of the stress state even if the particular weighting constants of the material under test are not known. It has also been found that the elastic constants of Al alloys are much more dependent on the particular alloy and the mechanical and heat treatment (Thompson et al. 1996; Schneider 1997, 2000).

Since the elastic material constants – Young's and shear modulus – are, respectively, influenced by the temperature and texture of the parts, it follows that the velocities of ultrasonic waves will also be impacted by temperature changes and by the degree and symmetry directions of the texture. The ultrasonic stress analysis requires one to distinguish between the influence of temperature and texture on the

one hand and of stress on the other. The temperature dependence of ultrasonic waves propagating in Fe and Al alloys is well known and hence easily taken into account. The change of the acousto-elastic constants with temperature can be neglected as long as the component's temperature is not changing more than about  $\pm 15^\circ$  around room temperature (Schneider 2000). In order to separate texture and stress influences, a lot of different approaches are published, e.g., by Thompson et al. (1996) and Schneider (2000), and the most suitable method is strongly dependent on the material, on the geometry of the component, and on the strength and symmetry of the texture. One approach to separate the influence of texture of rolled thin sheets and another approach to separate the texture of platelike components are discussed in the next chapter.

As seen from Eqs. (1, 2, and 3), the velocities of ultrasonic waves propagating along the same direction are influenced by all components of the stress state, and because of different weighting factors, the fraction of stress influence differs. Hence, it is recommended to apply two or even all three waves propagating in the component along the same path. The measuring quantity is the time of flight of the applied longitudinal or shear or guided wave (Becker et al. 2016).

The accuracy of ultrasonic stress analysis is influenced by the accuracy of the ultrasonic time-of-flight measurement by the accuracy associated with the acousto-elastic constants of the material and by the accuracy of the velocity or time-of-flight data representing the stress-free state. Using updated equipment, it is easy to get time-of-flight data with reproducibility better than  $\pm 1$  ns. By averaging the readings, a relative reproducibility in the range of 1 out of  $10^5$  is achievable.

The measuring error associated with the material dependent acousto-elastic constants or weighting factors is in the range of 5–10%. This error is directly influencing the final stress result (see Eqs. 1, 2, and 3).

The error associated with the value representing the stress free state shifts the zero point position along the stress axis. It is difficult to evaluate that error; hence the comparison with results of established techniques is regarded as helpful. In all cases, experienced by the authors up to now, there is an agreement of the results of ultrasonic techniques with the results of the hole-drilling or ring-core technique within  $\pm 50$  MPa in case of steel components and within  $\pm 30$  MPa in case of Al components. It can be stated that the accuracy of the ultrasonic stress analysis is in the same range as the accuracy of the established hole-drilling, ring-core, and sectioning techniques. The error associated with the separation or discrimination of the texture influence depends heavily on the applied approach.

---

## Applications

### Evaluation of a One-Axial Stress State

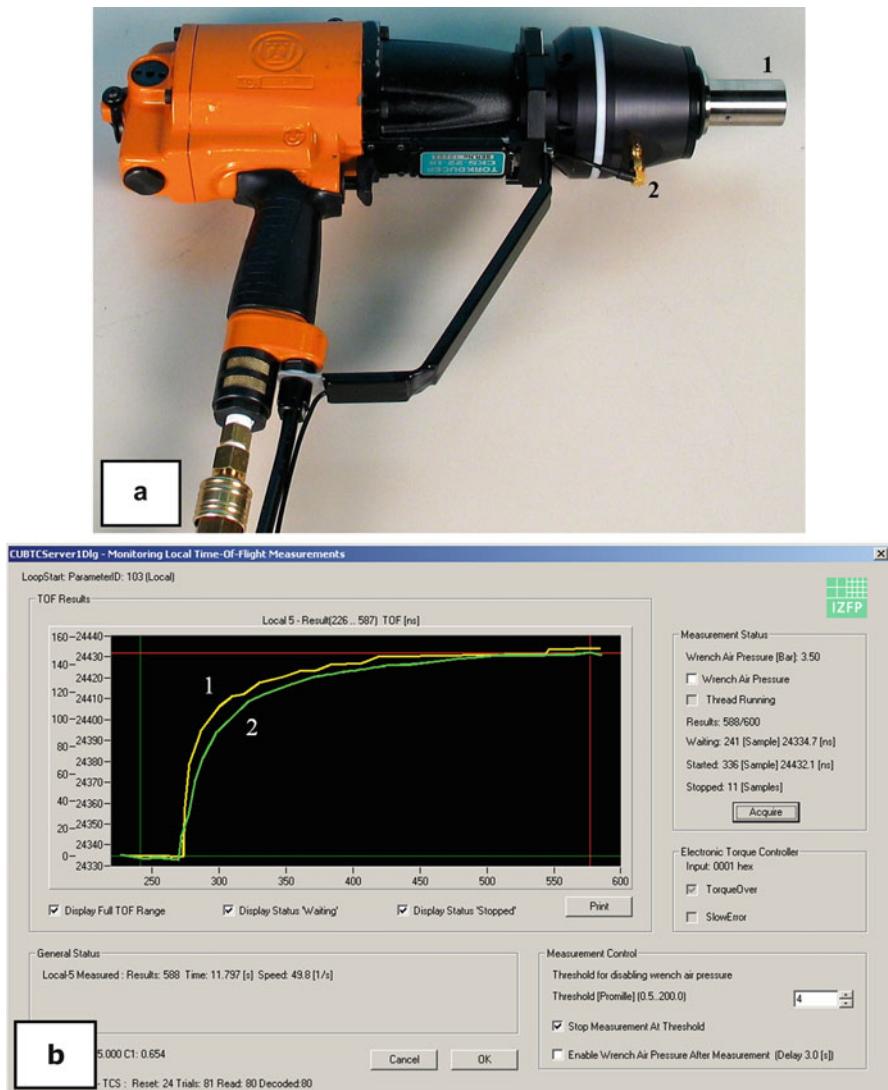
In order to evaluate one-axial stress states, a longitudinal wave propagating along the stress direction is frequently used. Since the atoms vibrate along the longitudinal wave propagation direction, the influence of the stress component parallel to the

propagation direction causes the most significant change of the velocity. It can be concluded from published values of the acousto-elastic constants that  $A/C$  is more than ten times higher in value than the value of  $B/C$  for all ferritic and austenitic steel grades. Hence, the application of a longitudinal wave propagating along the principal direction  $i$  enables the evaluation of the principal stress  $\sigma_i$  (Eq. 1). The influence of the principal stresses along the two directions perpendicular to the wave propagation direction can be neglected because of the small weighting factor. Nevertheless, for a general description of the stress situation, the variation of geometric boundary conditions of bolted joints has to be taken into account.

A very popular application of this approach is the evaluation of the longitudinal stress of bolts and the evaluation of the clamp load. The significant advantage of this technique is that there is no influence of the friction losses as in the case when the torque is taken as a measure. The ultrasonic technique evaluates the present stress state of the bolt. Systems on the market use either piezoelectric sensors glued on the surface of the bolt (Kibblewhite 1989) or a piezoelectric film sputtered on the surface (Intellifast 2009). Figure 1a shows a commercial pulse wrench with a usual socket in which an ultrasonic longitudinal wave transducer is incorporated. The electric signals are transmitted by induction principles from the rotating shaft to the outside of the tool (Fig. 1a – No. 2) and further to the ultrasonic front end. The measurement of the ultrasonic time of flight with a measuring rate of about 50 Hz is used to evaluate the increase of the stress or clamp load and to stop the process as soon as the requested value is reached. This particular system also takes the torque data. The display of the ultrasonic result (Fig. 1b; green No. 2) and the applied torque (Fig. 1b; yellow No. 1) enables a redundancy in the analysis of the bolt-tightening process and in the evaluation of the clamp load (Schneider and Herzer 2006).

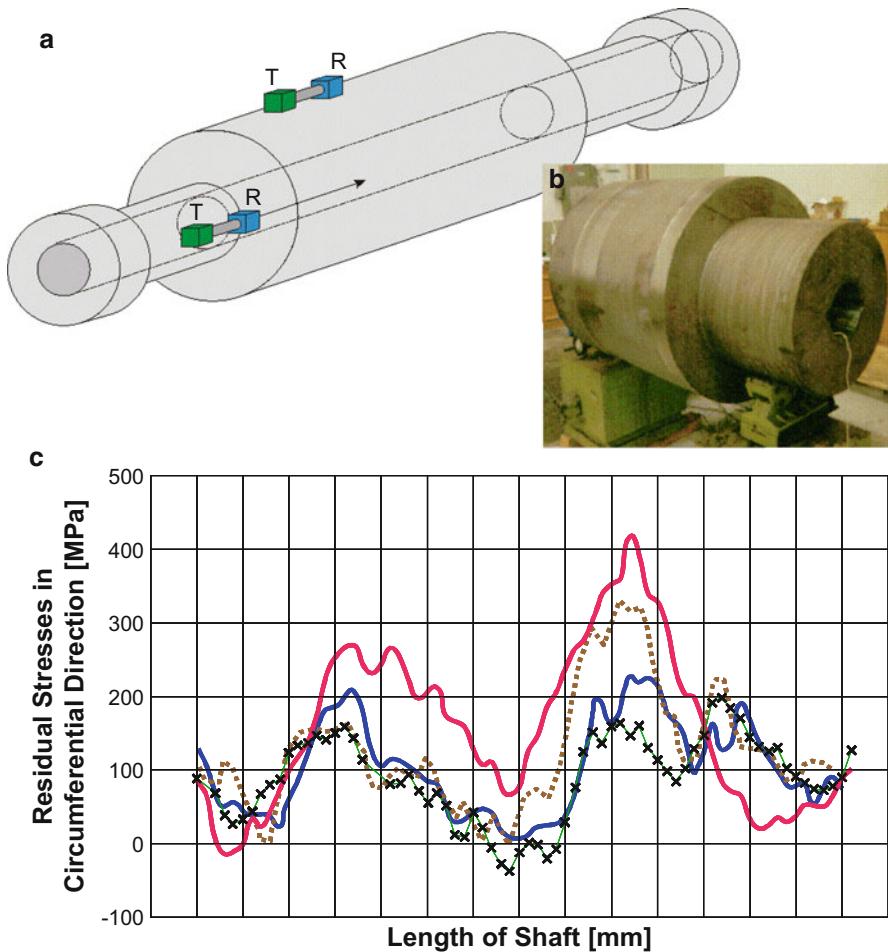
Another prominent case of a (quasi) one-axial stress state is the longitudinal stress in rails. It is mainly the straightening process which produces a tensile stress in the head of a new rail, and the traffic changes it into a three-axial compressive stress state. Track buckling may occur if the compressive stress becomes too high, e.g., during the summer. The group around Deputat and Szelazek (Brokowski and Deputat 1985; Szelazek 2001) developed a system to evaluate the longitudinal stress in the heads of new rails as well as of rails in the track. For that purpose, the longitudinal wave transducer is mounted to a wedge in order to generate the skimming longitudinal or critical angle longitudinal wave which propagates like a free wave in a near-surface layer. In order to separate the influences of changes of the microstructure along the measured part of the track, ultrasonic surface waves are also applied. This wave type is also generated by a longitudinal wave transducer mounted on a wedge with the critical angle for surface wave generation. Some systems have been developed and used by rail manufacturers and railway companies.

In some cases there is a three-axial stress state, but only one stress component is of interest. Forged turbine rotors are, for example, such components. It is the circumferential stress component which determines the major mechanical behavior and operational lifetime. The time of flight of an ultrasonic SH mode is measured. The wave propagates in a near-surface layer along the length direction of the rotor



**Fig. 1** Pulse wrench containing ultrasonic transducer in the socket (a – No. 1) and signal transmission unit (a – No. 2) for the on-line evaluation of the stress or clamp load (a). The screen in the lower part (b) displays the ultrasonic result (green No. 2) and the applied torque (yellow No. 1) of a bolt tightening process

and vibrates in the circumferential direction. The time of flight is mainly influenced by the circumferential (hoop) stress. In Fig. 2a and b, a typically forged part is sketched together with the positions of the ultrasonic transmitter-receiver set. In part (c) the profile of the hoop stress along a trace through the center bore hole is shown before and after different stress-relief treatments.



**Fig. 2** Sketch to visualize the ultrasonic hoop stress analysis on the outside surface layer and along the central bore hole of shafts and rotors (a). Part (b) shows a component investigated with the stress measurement system. The hoop stress along the length of a rotor in (c) is presented before (red/upper filled line) and after different mechanical and heat treatments to reduce the residual stress (dots, crosses and lower filled line)

### Evaluation of a Two-Axial Stress State

The stress state of the rims of railroad wheels can be regarded as two axial since the axial stress component is small. Railroad cargo wagons are braked by pressing braking shoes onto the running surface of the rims and the compressive stress of new rims is changed into a two axial stress state with a significant tensile stress component along the circumference of the wheel. This stress state of the rim may become critical by the accumulation of the influences of the braking forces. The commonly applied ultrasonic technique to evaluate the two axial stress state uses the

birefringence effect: The difference of Eqs. (2) and (3) results in the birefringence Eq. (4). The application of the birefringence equation is limited to components which do not exhibit a texture. Since the shear wave propagates the same path length, the velocities can be replaced by the times of flight  $t_{ij}$  and  $t_{ik}$ . Since  $v_T$  can be approximated by  $v_{ik}$ , Eq. (5a) and (5b) are obtained and used to evaluate the stress state of rims of railroad wheels.

$$\frac{v_{ij} - v_{ik}}{v_T} = \left( \frac{H}{K} - \frac{F}{K} \right) \cdot (\sigma_i - \sigma_k) \quad (4)$$

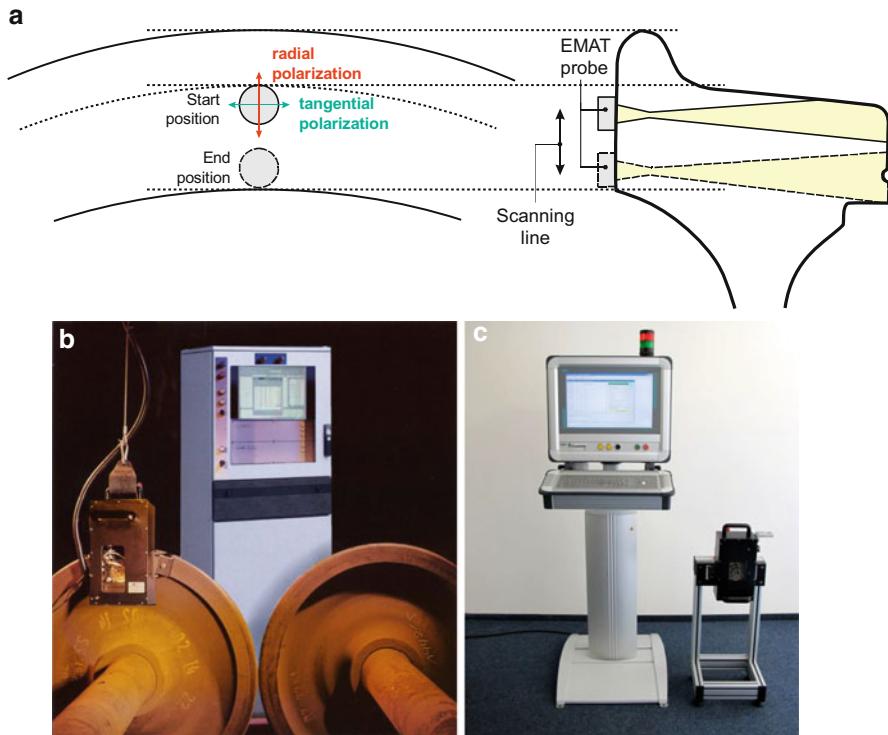
The relative difference of the times-of-flight of the shear wave vibrating parallel to the circumferential direction  $t_{\text{circumferential}}$  and of the shear wave vibrating parallel to the radial direction  $t_{\text{radial}}$  is a function of the difference of the principal stresses  $\sigma_{\text{circumferential}} - \sigma_{\text{radial}}$  along the two principal directions, respectively. It has been found that the principal stress along the radial direction  $\sigma_{\text{radial}}$  is small in value and does not change significantly under the influence of the braking. Hence, the difference of the two principal stresses  $\sigma_{\text{circumferential}} - \sigma_{\text{radial}}$  is mainly representing the circumferential stress component  $\sigma_{\text{circumferential}}$  and its change caused by the braking.

$$\frac{t_{\text{circumferential}} - t_{\text{radial}}}{t_{\text{radial}}} = \frac{\sigma_{\text{circumferential}} - \sigma_{\text{radial}}}{K_{\text{Wheel}}} \quad (5a)$$

$$\sigma_{\text{circumferential}} - \sigma_{\text{radial}} = K_{\text{Wheel}} \cdot \frac{t_{\text{circumferential}} - t_{\text{radial}}}{t_{\text{radial}}} \quad (5b)$$

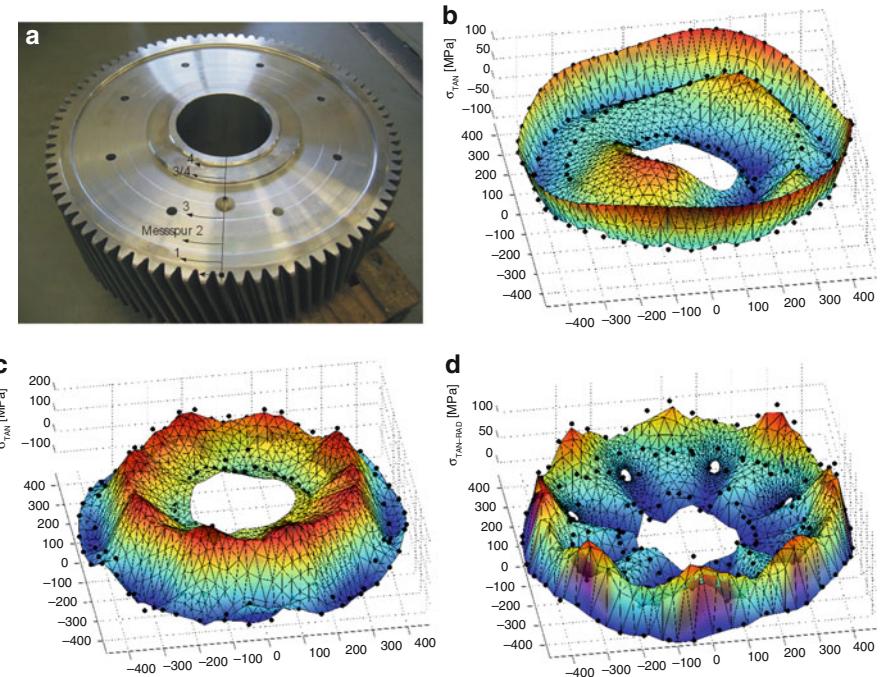
The proportionality is given by  $K_{\text{Wheel}}$ , a material-dependent acousto-elastic constant.  $K_{\text{Wheel}}$  is experimentally evaluated using a tensile test sample of the material of interest. Figure 3 visualizes under (a) the testing situation and the propagation and vibration directions of the applied shear wave. The first generation from 1992 (b) and the 2014 version of UER system (c) to evaluate the stress state of rims of railroad wheels are shown in the lower part of the figure. Over 50 automated systems are in worldwide use in workshops of railway companies and wheel manufacturers (Schneider et al. 1994; Herzer et al. 1994). The systems use an electromagnetic ultrasonic transducer (EMAT) to generate and to receive the polarized shear wave with the significant advantage that there is no viscous coupling medium needed between the sensor and the surface of the wheel. That facilitates the automated evaluation of the stress profile along the radial trace across the thickness of the wheel.

At about the same time Deputat and coworkers applied the birefringence effect also to evaluate the stress state of rims of railroad wheels, they developed DEBBI, a semiautomated system. Piezoelectric shear wave transducers are coupled to the wheel by the operator at each measuring position and are maneuvered along the radial trace (Deputat 1995; Jemec et al. 2008). Meanwhile, very similar systems with electromagnetic ultrasonic transducers were also developed by different groups. But from the viewpoint of ease of application, user friendliness, and grade of automatization, those systems cannot be compared with UER (Gilardoni et al. 2014; Jimenez et al. 2016).



**Fig. 3** Sketch of the measuring principle (a) and pictures of the first version (1992) (b) and the actual version (2014) of the UER system (c) to evaluate the stress state in rims of railroad wheels

The combined use of a longitudinal and a linearly polarized shear wave results in the visualization of the two individual stress values in platelike components or in the cross section of components like gear parts. Figure 4 displays a gear component (a) and results of the ultrasonic stress analysis in b, c, and d. The ultrasonic longitudinal and shear waves propagate in the thickness (axial direction) of the gear. The shear wave is polarized along the radial and along the circumferential direction, respectively. According to Eqs. (1, 2, and 3), the wave velocities are influenced by all three stress components. The stress component  $\sigma_i$  along the thickness direction is zero at the two surfaces and averages to zero after the integration over the cross section. In a first approximation, it can be assumed that the value of the stress component  $\sigma_i$  averaged along the thickness is small. The influence of the in-plane stresses  $\sigma_j$  and  $\sigma_k$  is weighted by  $B/C$  (Eq. 1). As can be seen in the appropriate table of the elastic and acousto-elastic material constants,  $B/C$  is small in case of ferritic steel grades (Schneider 2000). Hence, it is acceptable to assume that the time of flight of the longitudinal wave propagating along the thickness of the component is not influenced by the stress state. The time-of-flight data are used to evaluate the local thickness of the component. For that purpose the value of the longitudinal velocity for the particular material is used. Using the thickness and the times-of-flight of



**Fig. 4** Photo of a gear part (a) and profile of the hoop stress component of the gear part before (b) and after the teeth hardening (c) and the visualization of the difference of the two principal stresses  $\sigma_{tangential} - \sigma_{radial}$  (d)

the two shear waves, the shear wave velocities are calculated at each and every measuring point. An error of the assumed longitudinal wave velocity is not influencing the procedure since both shear wave velocities are influenced identically by this error. Now, using the two shear wave velocities, the Eqs. (2 and 3) can be used to evaluate the two principal stress components  $\sigma_j$  and  $\sigma_k$ . The weighting factors are taken from tables and  $v_T$  is experimentally evaluated using a representative sample of the steel grade. The inaccuracy of the individual stress value is dominated by the accordance of  $v_T$  with the shear wave velocity which would be measured in the stress-free gear. This inaccuracy determines the zero-stress point in the results of the stress analysis. But in case the local stress distribution is of interest or the change of the stress state after hardening or after stress relief treatments, this inaccuracy is of minor importance. The results show one of the most important advantages of the ultrasonic stress analysis which is the possibility to cover large areas (Schneider 2000).

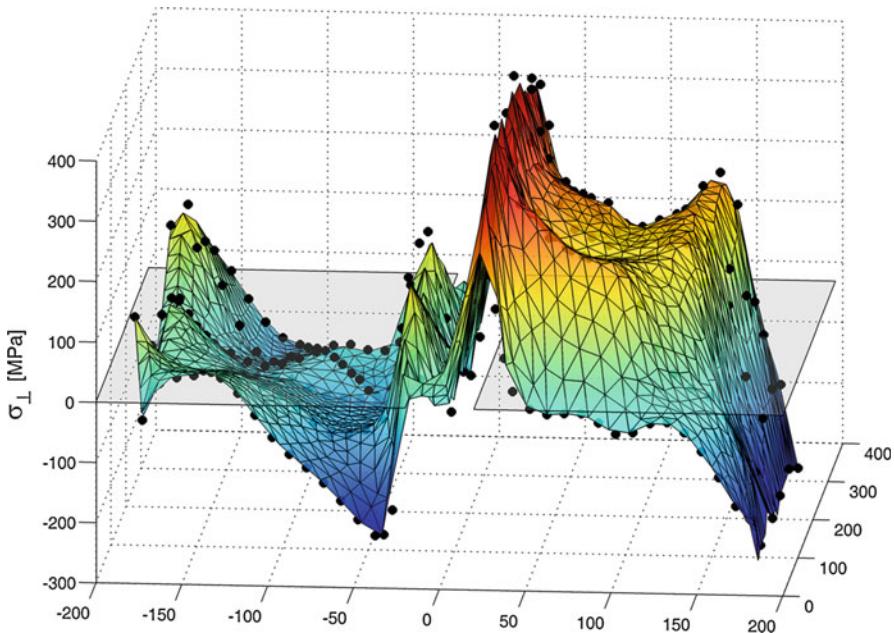
In order to characterize the two axial stress state in rolled sheets and plates, the symmetry of the rolling texture is exploited. As can be seen in, for example, (Schneider 2000) the influence of rolling texture on the velocity of a shear wave is the same when the directions of wave propagation and vibration are interchanged. The application of the birefringence evaluation results in Eq. (6).

$$\frac{v_{ij} - v_{ji}}{v_T} = \mu \cdot (\sigma_j - \sigma_i) \quad (6)$$

Equation (6) can easily be applied if EMAT sensors are used which generate the SHo plate wave mode. The Lamé constant  $\mu$  is identical with the shear modulus  $G$  of the isotropic material (Thompson et al. 1996; Schneider 2000). Hence, the relative difference of the velocities or times-of-flight is a direct measure for the difference of the two principal stress components acting along the rolling and the transverse direction in the plate. The assumption that the principal stress axes are parallel to the rolling and perpendicular to the transverse directions, respectively, is very often fulfilled. The visualization of the difference of two principal stresses along the measuring traces is found to be good enough to localize stress anomalies in rolled products as well as for the optimization of the rolling process and the stress relief treatment. In order to evaluate the individual values of the two principal stress components in the plane of rolled sheets or plates, the technique as described in the next paragraph can be applied.

If the two-dimensional stress state of a rolled component is of interest, the application of a longitudinal and a linearly polarized shear wave is of advantage. The waves propagate along the thickness direction  $i$ , the shear wave vibrates along the principal direction  $j$  and  $k$ , respectively (see Eqs. 2 and 3). In order to get a reliable value of the shear wave velocity which represents the stress-free state  $v_T$ , a sufficient number of measuring points have to be distributed over the components' surface. Applying the stress equilibrium conditions, the stress  $\sigma_j$  integrated over the thickness  $i$ ) – width ( $k$  direction) plane results in zero value as the stress  $\sigma_k$  integrated over the thickness – length ( $j$  direction) plane becomes zero. Hence, the mean value of all velocities of the shear wave vibrating along the  $j$  direction yields the velocity representing the stress free state, as the mean value of all velocities of the shear wave vibrating along the  $k$  direction yields the velocity representing the stress free state, respectively. If there is a rolling texture in the component, the value of  $v_T$  will be different depending on the vibration direction of the shear wave along the rolling direction and along the transverse direction of the rolled component. The two different values  $v_T$  characterize the influence of texture. Using the value of  $v_{T\text{ Rolling}}$  and of  $v_{T\text{ Transverse}}$ , the individual values for the two stress components  $\sigma_k = \sigma_{\text{Rolling}}$  and  $\sigma_j = \sigma_{\text{Transverse}}$  are calculated according to Eqs. (2 and 3). Taking the values of  $v_{T\text{ Rolling}}$  and of  $v_{T\text{ Transverse}}$  for the stress analysis, the influence of texture is separated. The only assumption to be made is a homogeneous texture along the length and the width direction of the component.

The two axial stress of a welded plate can be evaluated as described above. But in the area of the weld seam, the measurements are often not possible because of the weld reinforcement. Based on the stress profiles of the two plates welded together, the stress state of the weld seam can be evaluated using the stress equilibrium conditions for the welded plates or for a sufficiently expanded part of a welded structure. Figure 5 shows an example of such an analysis of the welding stress. The two plates welded together were cut from the same mother plate, having a strong texture. One plate was welded with the former rolling direction along the weld seam, the other with the rolling direction perpendicular to the plate.



**Fig. 5** Distribution of the stress component acting perpendicular to the weld seam in a thermo-mechanically rolled plate with a multilayer weld

### Evaluation of a Three-Axial Stress State

Since the velocities of ultrasonic longitudinal and shear waves propagating in the component along the same path length are influenced by the three principal stress components, it seems to be possible to evaluate the three individual stress values if the shape of the component enables the application of the sensors in such a way that the waves propagate along at least two principal stress directions. As mentioned earlier, the basic Eqs. (1, 2, and 3) are only valid if the directions of wave propagation and vibration are parallel to the principal directions of strain or stress. This condition limits the possibilities of the evaluation of a three axial stress state. The condition can be fulfilled using, e.g., thick plates. But the stress component along the thickness is not really influencing the mechanical behavior of thick plates. The distribution of the two stress components along the length and the width and their changes along the length, the width, and the thickness are much more influential. Therefore, there is no real need to evaluate the three individual stress components of plates at present.

There is some interest to get a tomography of the stress state of cylindrical components like rolls and rotors. In case of applying a transmitting sensor on the surface of a roll and the receiving sensor at the corresponding position of the circumference, the sound propagates along a tangential trace when the two sensors are close to each other. In this case the longitudinal wave will mainly be influenced by the tangential stress component. When the sensors are on both sides of

the diameter, the longitudinal wave will mainly be influenced by the radial stress component. Applying a radially vibrating shear wave in a similar way, the velocity will be influenced by the radial component of the stress state while propagating in the center part of the cylinder and by the tangential stress component when propagating in the outer parts. It is seen that the conditions of wave propagation and vibration parallel to the principal strain or stress directions is not fulfilled in sufficient manner in order to resolve the three axial stress state in a cylindrical component. Hence, it has to be stated that a three-axial stress state of a component cannot be evaluated by ultrasonic means only.

---

## Fields for Further Development of Ultrasonic Stress Evaluation

The presence of digitalized processes or process steps not only introduces several challenges for nondestructive testing but also offers opportunities to realize enhanced evaluation principles taking the results of further nondestructive testing methods into account (examples see ► Chap. 5, “[Guided Wave Testing](#),” ► Chap. 24, “[Micromagnetic Materials Characterization](#)”). In the field of nondestructive testing, two concepts should receive focus for future applications. The first is the so-called digital twin of a product, which is a digital representation of a component with information concerning all production steps and related states the component has already been subjected to. All data collected under this label are presented from the perspective of the product related to its ideal conditions and its ideal product lifetime or usage. A second but also important concept for nondestructive testing is the digital material data space, which is a digital representation of the material used for a component to combine all information about the microstructural and physical properties resulting from the production and usage of the component.

With these considerations in mind, tasks can be formulated that ultrasonic stress measurement will face in the near future. A central point is how data are stored within the measurement process. Although only the time of flight is relevant for the stress calculation, the whole process generates much more information such as A-images, temperatures, ultrasonic excitation, and receiver properties. If this information is additionally stored and connected to existing material and process data, the evaluation with respect to a product’s entire life cycle is possible with minimum effort. For example, the ultrasonic signals can transport information about the microstructure of the material or the geometric properties of the component, neither of which can be adequately represented by only the time-of-flight value at a certain point.

There are numerous expected benefits for ultrasonic stress measurement resultant from the digital integration in industrial application. The following innovations would be the most promising next steps for the authors to overcome known restrictions.

The accessibility of digital design data as well as values of the material properties, tolerances, and other related information resulting from quality assurance processes would enable the improvement of the selection of measurement positions and the separation of geometric changes from stress effects. Hence, this approach could lead to improved accuracy in evaluated stress values. In addition, the existing boundary

conditions could be related to changes in the shape of ultrasonic signals with the objective of separating scattering and attenuation effects from relevant stress-related time-of-flight changes. Another example is the use of three dimensional models of the component to calculate its behavior under different superimposed loads, leading to the estimations of the local stress values along the sound path of the ultrasonic wave. Hereby, different load scenarios and their influence on the ultrasonic wave could be calculated by finite element analysis. These supportive applications of the calculations in combination with known relations of the stress-equilibrium can also be used to allow the interpolation of stress values at positions not accessible for ultrasonic stress evaluation. There are much more scenarios conceivable as to how the digitalized processes and “digital twins” can improve ultrasonic stress measurements leading to the long-term objective of the highly individual lifetime monitoring of single safety-relevant components.

---

## Summary

The physical basics of the ultrasonic techniques to evaluate stress states of components are very well documented and understood. The influence of any stress state on ultrasonic velocities can be calculated; the best choice of ultrasonic wave combination in order to minimize texture and microstructure effects can be made. The state of electrical and software engineering yields semi- and fully automated systems to measure the time of flight of ultrasonic waves with sufficient resolution and accuracy and to evaluate the individual stress component or to characterize the stress state of a component by mapping the local changes of stress states. A variety of commercial sensors are available to meet the requirements concerning ultrasonic wave mode and probe size and shape on one side and the stress state of interest on the other side; design concepts and broad experience to manufacture specific electromagnetic ultrasonic transducers (EMATs) and piezoelectric sensor arrays is available. Although the ultrasonic techniques can be an efficient tool in the quality assurances concepts of industrial partners, they are not widely used. There is no Just-Take-and-Apply-Technique available; sensor and measuring technique have to be adapted to each particular case of interest. The efforts to integrate adequate ultrasonic methods for stress evaluation in complex systems will grow with the increase of digitalization of production lifecycle management. Only a very small number of researchers and engineers developed systems which are applied in an industrial environment for very specific cases of application.

## Evaluation of Clamp Load of Bolts

1. Widely used in the automotive industry are commercial ultrasonic thickness gauges to measure the length of each bolt before and after tightening. Calibrations are used to determine the length corresponding to the requested clamp load and a comparison with the final length of the bolt, to be done by the operator, characterizes the quality of the joint.

2. Automated systems are on the market to evaluate the clamp load using calibration data taken for a large variety of bolt types. Taking data before and after the tightening process, the clamp load or the elongation or the longitudinal stress can be evaluated.
3. Automated systems are used to evaluate the clamp load of bolts in wind energy plants and airplane parts. Since the piezoelectric element is glued or sputtered on each of the bolts, the systems are mainly applied on very special and critical joints. The systems are applied during the tightening process as well as used for the periodical check of the clamp load.
4. The incorporation of a commercial ultrasonic sensor into the socket of a commercial bolt tightening system allows an evaluation of the clamp load during the tightening process and the optional recording of the applied torque yields a redundant quality control. This on-line measurement is also used to stop the process as soon as the requested clamp load is reached. These automated systems are used on special automotive parts and on large bolts in parts of the turbine housing.

## Evaluation of Stress States of Railroad Wheels

There are four European manufacturers known to the authors using the same physical background. The systems differ mainly in the grade of automation and the presentation of the results of the stress analysis. The most widely used system uses an EMAT transducer which is automatically moved along a radial trace over the thickness of the rim. The evaluated stress profile is compared with the mean stress and max stress values, given by the railroad authorities for each type of wheel. Green or red light indicates to the operator whether the stress state of the wheel allows its further use or not.

## Other Applications

Prototype systems developed by particular groups of researchers are applied to evaluate stress states in different parts and components. The applications are made in the framework of measuring services commissioned by industrial partners. Only a very limited number of applications have been published. Results of the following applications are known to the authors:

- Evaluation of the longitudinal stress in new rails and in rails on the track.
- Evaluation of the surface near hoop stress along the length of rolls, hardened rolls, and turbine rotors. Measurements are done on the surface as well as along the center bore hole. The main purpose was to support the appropriate stress relief treatment of the component.
- Evaluation of the two principal stress components along the length and along the width of rolled steel and Al-sheets to enable a feedback to the rolling conditions and the mechanical and heat treatment of the sheets.

- Evaluation of the welding stress state of steel and Al-sheets and plates to detect hot spots and anomalies of the stress distribution and to enable a feed back to the welding parameters.
- Evaluation of the stress distribution in Al-heavy plates as a tool to optimize the heat treatment in order to generate a homogeneous stress distribution.

---

## References

- Becker M, Gross N, Herzer R (2016) Determination of preload in bolts by ultrasound without referencing in unloaded state. In: Conference proceedings: 19th world conference of nondestructive testing
- Brokowski A, Depurat J (1985) Ultrasonic measurements of residual stresses in rails. In: Conference proceedings: 11th world conference of nondestructive testing, pp 592–598
- Depurat J (1995) DEBBIE the fast nondestructive measurement of stress for in service railway wheel inspection. DEBRO UMS, Akademicka 3, 02-038 Warsaw
- Egle DM, Bray DE (1976) Measurement of acoustoelastic and third order elastic constants for rail steel. *J Acoust Soc Am* 60(3):741–744
- Gilardoni C, Gherbin M, Carboni M, Gianneo A (2014) High-performance methodology for residual stress measurement in railway wheels. In: Conference proceedings: 11th European conference of nondestructive testing
- Herzer R, Frotscher H, Schillo K, Bruche D, Schneider E (1994) Ultrasonic set-up to characterize stress states in rims of railroad wheels. In: Green RE, Kozaczek KJ, Ruud CO (eds) Nondestructive characterization of materials VI. Springer, Boston, pp 699–706
- Hughes DS, Kelly JL (1953) Second order elastic deformation of solids. *Phys Rev* 94(5):1145–1149. <https://doi.org/10.1103/PhysRev921145>
- Intellifast GmbH (2009) Bolting failure-free technology on the offshore wind industry. In: Power & energy & solutions Europe, pp 54–57. Available via PES Europe. <http://cdn.pes.eu.com/assets/misc/corporate-focus-intellifast-gmbhpdf-35.pdf>. Accessed 29 Mar 2018
- Jemec V, Grum J, Bozicko S (2008) Ultrasonic measurement of hoop stress in the rim of monoblock railroad wheel. In: 38th international CNdT conference Defektoskopie 2008. Conference Proceedings, Brno, pp 269–274
- Jimenez JA, Garcia V, Boyero C (2016) Handheld solution for measurement of residual stresses on railway wheels using EMATs. In: Conference proceedings: 19th world conference of nondestructive testing
- Kibblewhite IE (1989) Ultrasonic load indicating member, apparatus and method. Patent number 4899591. Type: Grant. Assignee: SPS Technology, Inc.
- Murnaghan FD (1951) Finite deformation of an elastic solid. Wiley, New York
- Schneider E (1997) Ultrasonic techniques. In: Hauk V (ed) Structural and residual stress analysis by nondestructive methods. Elsevier, Amsterdam, pp 522–563
- Schneider E (2000) Untersuchung der materialspezifischen Einflüsse und verfahrenstechnische Entwicklungen der Ultraschallverfahren zur Spannungsanalyse an Bauteilen. Fraunhofer IRB, Stuttgart
- Schneider E, Herzer R (2006) Ultraschall-System zur on-line Bestimmung der Schraubenvorspannkraft und zur Schraubersteuerung. *ZfP-Zeitung* 100:40–46
- Schneider E, Herzer R, Bruche D, Frotscher H (1994) Ultrasonic characterization of stress states in rims of railroad wheels. In: Green RE, Kozaczek KJ, Ruud CO (eds) Nondestructive characterization of materials VI. Springer, Boston, pp 383–390
- Szelazek J (2001) Postepy W Ultradzwiekowych Badaniach Naprezen. Praca Habilitacyjna ISSN0208–5658
- Thompson RB, Lu WY, Clark AV Jr (1996) Ultrasonic methods. In: Lu J (ed) Handbook of experimental mechanics. The Fairmont Press Inc., Lilburn, pp 149–178



# Nonlinear Acoustics

8

Younho Cho and Weibin Li

## Contents

Definition and Historical Background .....	252
Earlier Efforts .....	253
Higher Harmonic Generation .....	253
Nonlinear Wave Modulation Spectroscopy .....	254
Shift of Resonance Frequency .....	255
Recent Trends .....	256
Slow Dynamics .....	256
Subharmonic Phased Array Technique .....	256
Nonlinear Laser Ultrasonic Technique .....	258
Nonlinear Ultrasonic NDT .....	258
Nonlinear Ultrasonic Bulk Waves NDT .....	259
Nonlinear Ultrasonic Guided Waves NDT .....	260
Concluding Remarks .....	265
References .....	267

### Abstract

Early detection and continuous tracking of material micro-damages have been one of the most demanding techniques in industries. Due to the sensitivity of acoustic nonlinearity to micro-damage, the nonlinear ultrasonic technique has been explored as a promising tool for early detection of micro-damages. In this chapter, we briefly introduce the earlier efforts and recent development of the nonlinear acoustics and their applications for nondestructive testing and

---

Y. Cho (✉)

School of Mechanical Engineering, Pusan National University, Busan, South Korea  
e-mail: [mechcyh@pusan.ac.kr](mailto:mechcyh@pusan.ac.kr)

W. Li

School of Aerospace Engineering, Xiamen University, Xiamen, China  
e-mail: [liweibin@xmu.edu.cn](mailto:liweibin@xmu.edu.cn)

evaluation (NDT& E). Some advanced techniques based on measure of nonlinear acoustics for NDT& E are also introduced as potential and attractive means.

---

## Definition and Historical Background

Nonlinear acoustics is a branch of physics and acoustics dealing with acoustic waves propagation in either fluid or solid media with nonlinear response. Generally, in the case of acoustic waves with sufficient amplitudes in the media, the wave motion should be governed by nonlinear wave equations for the reason that linearization is no longer possible in this situation. In solid materials, grain boundaries, lattice anharmonicity, energy absorption, imperfect interfaces etc. can be the sources for arising nonlinearity in the solid media (Landau and Lifshitz 1970; Campos-Pozuelo et al. 2006; Kim et al. 2006b). Due to material nonlinearity, acoustic waveform propagation in the media can distort. The acoustic nonlinear responses include creating accompanying harmonics, multiplication of waves of different frequencies, and, under resonance conditions, changes in resonance frequencies as a function of drive amplitude (Van Den Abeele et al. 2000a).

The use of nonlinear acoustics in solid media can be significant in nondestructive testing (NDT) for the interrogation of micro-damages in materials at an early stage. In other words, the sensitivity of nonlinear acoustic methods for the detection or evaluation of material damage is far higher than that of linear ones. Thus, acoustic nonlinear responses in solid media can be used as a promising nondestructive testing method in solid media for micro-damage detection. Acoustic nonlinearity is usually characterized by a quantitative parameter that quantifies the amount by which an ultrasonic wave is distorted as the acoustic wave travels through the specimen. More detailed definition of nonlinear parameter for various types of ultrasonic waves in solid media will be introduced later in this chapter. The principle of nonlinear ultrasonic NDT is based on the fact that the level of acoustic nonlinearity in the materials containing damage is more than those in materials with no damage.

For acoustic waves propagation in media with nonlinearity (micro-cracks or damage), various nonlinear phenomena will be generated. Thus, different nonlinear acoustic techniques are developed based on the measure of these phenomena. For acoustic wave of single frequency propagating through a nonlinear medium, new components of higher and lower frequency can be generated in the material. Higher-frequency waves must have frequencies that are integer multipliers of the input wave frequency, which is usually called higher harmonics generation technique (Hikata and Elbaum 1966). Under certain conditions the generated wave frequency can be half of the input wave frequency, which corresponds to sub-harmonics generation technique (Chomas et al. 2002). For two or more acoustic waves with different frequencies propagating in the nonlinear media, the interactions of these waves with material micro-damages can produce nonlinear components with other frequencies. These new components with either a sum or difference of frequencies can be used to evaluate or detect the material nonlinearity (Kuvshinov et al. 2013). Monitoring the resonance frequency shift with the increase of excitation amplitude is another nonlinear acoustic-based technique

(Van Den Abeele et al. 2000b; Muller et al. 2005). Nonlinear wave modulation spectroscopy is also an interesting technique that has been widely used to evaluate material nonlinearity. The technique utilizes a continuous high-frequency probe wave and a low-frequency vibration or pump wave. The pump wave is generally a given resonance mode of the tested sample (Kober and Prevorovsky 2014).

Acoustic nonlinearity can be used for characterization of microstructural evolution and detection of material damage. Initially, micro-damaged materials show progressively enhanced features of nonlinear acoustic response. Classical nonlinear acoustic theory has explained the nonlinear behavior in the material with nonlinearity at the atomic/molecular scale very well. The enhanced nonlinear response arising from the complex compliance of local or volumetric damage dominates the nonlinear sources. However, there are two issues that need to be clarified:

1. The mechanism of acoustic nonlinear response in solid media is far more complicated and not yet well understood. In addition, even for the classic quadratic nonlinearity, the measured acoustic nonlinearity should also be carefully clarified (Qu et al. 2011).
2. For different types of damage in solid media, the corresponding nonlinear mechanisms are also different and the nonlinear behavior of acoustic waves in solid media with cracks or damage may induce coupling phenomena with classical nonlinearity, as well as hysteresis and discrete memory (Van Den Abeele et al. 2000a). Thus, quantitative characterization of material nonlinearity by acoustic waves has not yet been demonstrated.

---

## Earlier Efforts

The use of nonlinear acoustic waves for nondestructive testing is receiving increasing attention for the high sensitivity of the approach (Donskoy et al. 2001). Earlier efforts on the theory and applications of nonlinear acoustic have been reviewed in text books, chapters, as well as review papers. M.F. Hamilton and D. T. Blackstock edited one book about nonlinear acoustic in 1998 (Hamilton and Blackstock 1998). In this book, a systematic introduction of nonlinear acoustics is provided. Considering the rapid development in nonlinear acoustic techniques, some researchers reviewed the advance of nonlinear ultrasonic techniques for nondestructive assessment of micro-damage in solid media (Jhang and Kim 1999; Matlack et al. 2015; Li et al. 2017). Recently, J. Rushchitsky edited a book which provides a coherent treatment of the theory of propagating nonlinear elastic waves in solid media (Rushchitsky 2014). The principles of some typical nonlinear ultrasonic techniques are addressed in this chapter.

## Higher Harmonic Generation

Higher harmonic generation is one of the typical nonlinear phenomena for acoustic waves in nonlinear media, and it is viewed as the classical acoustic nonlinear

phenomenon. Physically, the phenomenon of higher harmonic generation is related to nonlinearity in the elastic behavior of material (Goldberg 1956). To date, a number of investigators have applied this nonlinear ultrasonic technique to assess fatigue damage in different materials. The correlation between dislocation density levels within the fatigued material and increase in acoustic nonlinearity has been reported (Nagy 1998; Jhang 2000). Many researches have carried out measurements of acoustic nonlinearity versus material degradation in different kinds of material (Nagy 1998; Jhang 2000; Cantrell and Yost 2001; Kim et al. 2006a; Li et al. 2012a). The trend of acoustic nonlinear response in the specimen with improved material properties by heat treatment was also researched (Li et al. 2013). The second harmonic generation of surface wave was discussed, and experiments on SAW (surface acoustic wave) harmonic generation were first reported in aluminum and steel (Rischbieter 1967; Herrmann et al. 2006a). Vella discussed the nonlinear interaction of two collinear SAWs, where the rigorous theory of thermoelasticity was used to derive exact expressions for the nonlinear force and stress fields (Padmore and Stegeman 1976). Considering that second harmonic generation accompanies more structural information and the distinct advantages of guided wave techniques, nonlinear ultrasonic guided waves draw significant attention from the NDE community for material characterization and micro damage detection. An investigation of second harmonic generation of guided waves in isotropic plates has been first reported by Deng (1999). de Lima and Hamilton (2003) investigated the second harmonics field of elastic wave propagation in an isotropic plate. Srivastava et al. reported the possibility of existence of antisymmetric or symmetric Lamb wave modes at higher harmonics (Srivastava and di Scalea 2009). Higher harmonic generation of various guided waves has been used to evaluate material nonlinearity either in plates or tube-like structures as well as composites (Pruell et al. 2007; Liu et al. 2013; Li et al. 2012).

## Nonlinear Wave Modulation Spectroscopy

Nonlinear wave modulation spectroscopy is another nonlinear acoustic technique for evaluation of material nonlinearity. The technique is based on the interactions of two waves with probe frequency and pump frequency manifesting themselves as side-band components in the frequency spectra of the received signal. The nonlinear mechanism has been illustrated through the behavior ensued by the existence of a crack in a sample by Sutin (Solodov 1998). In addition, vibro-acoustic modulation techniques were formerly conceived for the detection of localized defects such as cracks in structural parts. The Nonlinear response of acoustic waves modulation spectroscopy caused by various nonlinearity sources was also reviewed (Van Den Abeele 2000). In cases of localized defects such as cracks, the modulation arises from the nonlinear stiffness that ensues from the interfacial contact. In such cases, the behavior can even be chaotic (Solodov and Korshak 2002). Compared to the higher harmonic generation technique, the modulation technique offers some advantages (Donskoy and Sutin 1998). First, higher harmonic generation requires a homogeneous travelling path to take advantage of the cumulative effect – thus this

is difficult to fulfill in the presence of reflecting boundaries and other structural inhomogeneities. Second, high voltages are needed, which frequently add some nonlinear background signal and that may affect the sensitivity of the technique. Recently, nonlinear impact resonance acoustic spectroscopy (NIRAS) was proposed based on the conventional nonlinear wave modulation technique. It has been shown to be highly sensitive to defects, especially to small cracks in materials (Klepa et al. 2012; Eiras et al. 2014; Hilloulain et al. 2014). Wave mixing can also be considered as a special case of nonlinear wave modulation spectroscopy, which, due to the cross-interactions of two acoustic waves at different frequencies, happens mainly in certain mixing zones in the specimen (Croxford et al. 2009; Demcenko et al. 2012).

## Shift of Resonance Frequency

Nonlinear resonance techniques monitor the resonance frequency shift and attenuation variations with increasing amplitude of the excitation. The resonance frequency and attenuation are determined for different excitation levels. The material nonlinearity is manifested as a downward resonance peak shift and a decrease of the quality factor ( $Q$ ) – inverse of attenuation – with increasing excitation amplitude. Therefore, from the downward resonance frequency shift, the third-order elastic nonlinear term can be obtained as

$$\frac{f - f_o}{f_o} \approx \frac{\delta}{2} \cdot \varepsilon^2 \quad (1)$$

where  $\varepsilon$  is strain,  $f_o$  is the resonance frequency in the linear strain range,  $f$  is the frequency in the nonlinear strain range, and  $\delta$  is a measure of material nonlinear parameter. In practice,  $f_o$  is the resonance frequency for the lowest excitation level. The technique can be used to provide insight into the ultimate stress in brittle materials such as concrete and the yield stress for ductile materials (Zaremba et al. 1989). However, experimental evidence in polycrystalline solids and rocks revealed a linear dependence of the resonance frequency and attenuation shifts with strain amplitude along with an unexpectedly high third harmonic amplitude (Read 1940; Guyer et al. 1995; Guyer and Johnson 1999; Johnson and Rasolofosaon 1996). These observations do not align with the classical nonlinear behavior and were associated with hysteresis in the stress-strain relationship. Such behavior was found to be the characteristic of materials with defects at the mesoscale level: rocks, concrete, soil, cracked materials, etc., which are collectively termed nonlinear mesoscopic elastic materials, NMEM (Guyer and Johnson 1999). By including hysteresis in modeling resonance experiments (Nazarov et al. 2003), it was demonstrated that the resonance frequency shift is proportional to the strain amplitude ( $\Delta\varepsilon$ ), so that

$$\frac{\Delta f}{f_o} = \alpha_f \cdot \Delta\varepsilon \quad (2)$$

along with a linear decrease of attenuation as

$$\frac{1}{Q} - \frac{1}{Q_o} = \alpha_Q \cdot \Delta\varepsilon \quad (3)$$

where  $Q$  is the quality factor (inverse of attenuation) and  $Q_o$  is the quality factor obtained in linear strain regime. These dependences (Eqs. 2 and 3) may be of a higher order if the characteristics of the hysteretic function change (Pecorari and Mendelsohn 2014). The parameters  $\alpha_f$  and  $\alpha_Q$  quantify the extent of hysteresis and are presumed to have the same physical origins (Johnson and Sutin 2005).

## Recent Trends

### Slow Dynamics

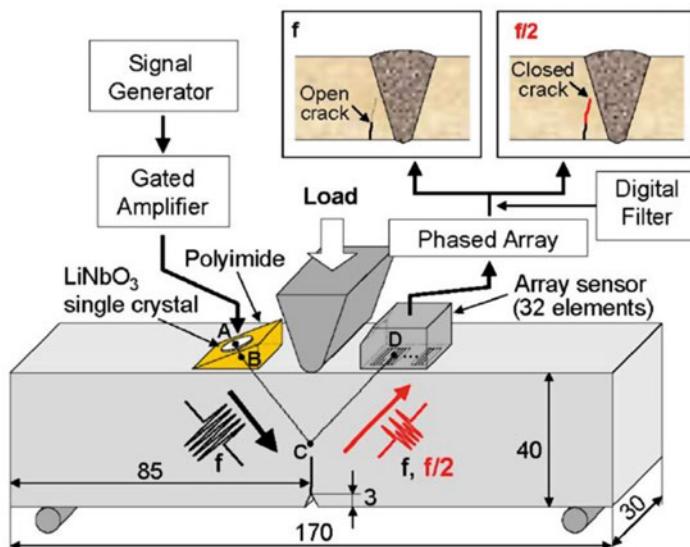
Slow dynamics (SD) is a novel nonlinear acoustic method, which is based on monitoring changes in the thermodynamic state of the material. This may be done, for example, by recording the resonant frequency for a specific resonance mode of an object and obtaining the change in the sound wave speed. One way to monitor SD is to probe the change in resonant frequency with a low-amplitude acoustic wave before and, again, after some disturbance of the thermodynamic equilibrium state. Rheological models were used to simulate the phenomenological behavior of slow dynamics (Bentahar et al. 2006; Favrie et al. 2015; Guyer et al. 1998; Nazarov and Radostin 2015; Ten Gate and Shankl 1996). However, although the nonlinear hysteresis and relaxation can be fairly well represented by these models, the underlying mechanisms may be different in different materials (Nazarov and Radostin 2015). Indeed, the physical origins of slow dynamics are still not very well understood. On the other hand, there seems to exist a strain amplitude threshold while at low strain amplitudes – say below  $\sim 10^{-7}$  the material exhibits nonlinear classical behavior. Beyond this threshold amplitude, hysteresis is activated, which is accompanied by a slow dynamic recovery; that is, the material enters a nonequilibrium or nonclassical regime. Slow dynamic effects appear to be related to the damage features such as micro-cracking. Various resonance-based techniques have been developed to assess the mechanical integrity of various materials (Van Den Abeele et al. 2000b). Slow dynamic effects coexist during dynamic excitation (Johnson and Sutin 2005), so that the measurement of the hysteretic parameters may be affected by slow dynamic effects. Such an effect can be minimized by increasing the time lapse between consecutive acquisitions.

### Subharmonic Phased Array Technique

To enhance the selectivity in the nonlinear ultrasonic response induced by micro-cracks, recently the subharmonic phased array technique was developed

(Ohara et al. 2007, 2009; Sugawara et al. 2015). It combines the sensitivity of nonlinear ultrasonics and the high power input of the phased array approach. Ohara et al. (2007) proposed a practical closed-crack imaging apparatus, called the subharmonic phased array for crack evaluation (SPACE) on the basis of subharmonic generation by short bursts and phased array algorithm with frequency filtering.

The experimental configuration of SPACE is shown in Fig. 1; a LiNbO<sub>3</sub> single-crystal transmitter with a polyimide wedge was used to generate intense ultrasound and a phased array sensor system was used as a receiver to concentrate on reception. When high-energy ultrasonic excitation is used, scattering of fundamental and subharmonic waves occur at the open and closed parts of the crack, respectively. The scattered waves received by the array sensor are converted to digital signals. Subsequently, they are digitally filtered at fundamental and subharmonic frequencies. After their phase shift following the delay law, they are added. Finally, the root-mean-square value is calculated as intensity at a focal point. This process is repeated over a scan area with incremental steps to create images. The fundamental and subharmonic images obtained can indicate the open and closed parts of cracks, respectively. Researchers used the SPACE technique to evaluate closed fatigue cracks and stress corrosion cracks. The measurement accuracy of SPACE for such micro-cracks is yet to be determined. It has been demonstrated that SPACE is very useful for correcting the underestimation of crack depths.



**Fig. 1** Experimental configuration of SPACE (From Ohara et al. (2007))

## Nonlinear Laser Ultrasonic Technique

Traditionally, contact methods based on piezoelectric transducers and/or capacitive probes have been used to monitor the generation of harmonic energy in a material substrate. These approaches suffer from a number of problems, however, that place restrictions or limits on measurement capabilities. There are often special requirements for specimen preparation, for example, that may require optically flat and parallel surfaces, or requirements on specimen conductivity. In addition, measurements are typically limited to a single measurement location, due to the need for hard-bonding of transducers or restrictions imposed by the placement of the probes. The spatial resolution capabilities of a typical measurement are also very coarse in nature, and involve an averaged or integrated signal over the entire receiver area, which can impact the quality of the measurements. An alternative method for detecting and characterizing the harmonic ultrasonic field in a material involves the use of laser interferometry techniques (Scruby and Drain 1990; Jia and de Billy 1992; Moreau 1995; Hurley and Fortunko 1997; Stratoudaki et al. 2011). Laser interferometry has long been used for dynamic motion measurements and it offers several advantages for making nonlinear ultrasonic measurements. Because coherent light is used as the probe, measurements can be made in a noncontact, remote, and nonintrusive manner. High spatial resolutions are also possible (1–10 microns) without reductions in measurement sensitivity. Interferometric measurements are also directly related to the optical wavelength used, which provides an absolute measure of the ultrasonic displacement levels. They also have a truly broadband frequency response, which is difficult to achieve with piezoelectric probes. And, finally, by raster-scanning the laser beam position with respect to the material surface, a high-resolution image of the harmonic (and fundamental) ultrasonic displacement field(s) can be created.

Recently, laser techniques for generation of nonlinear SAW pulses were developed resulting in the observation of strong nonlinear effects, such as the formation of shock fronts and drastic changes of the pulse shape and duration (Jia and de Billy 1992). It was demonstrated that a nonlinear compression as well as an extension of a wave pulse may take place depending on the nonlinear acoustic parameters (Hess et al. 2014; Liu et al. 2013; Liu et al. 2014). A noncontact testing with laser-generated ultrasonic transmitter is a very attractive technique in surface wave inspection. It can be used for online inspection and structural health monitoring, where contact methods with embedded sparse sensors cannot be applied such as are the cases of high temperature applications.

---

## Nonlinear Ultrasonic NDT

In this chapter, we view the second harmonic generation as the typical example to illustrate the nonlinear ultrasonic NDT methods. Thus, the physical meaning of nonlinear ultrasonic technique in this section is the measure of second harmonic waves for nondestructive testing.

## Nonlinear Ultrasonic Bulk Waves NDT

To detect material nonlinearities, a single frequency ultrasonic wave is launched into the specimen, and the signal responses from the ultrasonic wave as a function of propagation distance are received. The ultrasonic wave is distorted due to material nonlinearity, and consequently, higher harmonics are generated (Li et al. 2012). Thus, the received signal is composed of not only the fundamental frequency wave but also second or higher harmonic frequency waves. The measurement of harmonic generation for microstructural characterization is typically aimed at determining the value of the nonlinear acoustic parameter. The phenomenon of second harmonic generation is related to nonlinearity in the elastic behavior of material when the relationship between the one dimensional stress and strain is nonlinear (Cantrell and Yost 2001):

$$\sigma = E\varepsilon \left( 1 - \frac{1}{2}\beta\varepsilon \right). \quad (4)$$

Substituting Eq. (4) into the equation of motion of a solid element

$$\rho \frac{\partial^2 u}{\partial t^2} = \frac{\partial \sigma}{\partial x}, \quad (5)$$

yields the nonlinear wave equation as follows:

$$\rho \frac{\partial^2 u}{\partial t^2} = E \frac{\partial^2 u}{\partial x^2} - E\beta \frac{\partial u}{\partial x} \frac{\partial^2 u}{\partial x^2}, \quad (6)$$

where  $E$  is Young's modulus,  $\beta$  is the nonlinear coefficient,  $u$  is the displacement,  $x$  is a coordinate value in the wave propagation direction,  $\sigma$  is the longitudinal stress, and  $\rho$  is the mass density, respectively.

Since the amplitude of the second harmonic wave is much lower than that of the fundamental wave, a perturbation method can be used to solve this nonlinear equation. The displacement  $u$  is assumed to be of the form

$$u = u_1 + u_2. \quad (7)$$

Using the perturbation method, the following two equations are obtained:

$$\rho \frac{\partial^2 u_1}{\partial t^2} = E \frac{\partial^2 u_1}{\partial x^2}, \quad (8)$$

and

$$\rho \frac{\partial^2 u_2}{\partial t^2} = E \frac{\partial^2 u_2}{\partial x^2} - E\beta \frac{\partial u_1}{\partial x} \frac{\partial^2 u_1}{\partial x^2}. \quad (9)$$

By choosing  $u_1$  as a sinusoidal wave of a single frequency

$$u_1 = A_1 \sin(kx - wt), \quad (10)$$

the second-order solution can be obtained as

$$u_2 = A_2 \cos 2(kx - wt), \quad (11)$$

where  $A_2 = A_1^2 k^2 \beta x / 8$ ,  $A_1$  is the amplitude of the fundamental wave and  $A_2$  is the amplitude of the second harmonic, while  $k$  is the wave number. The acoustic nonlinear parameter  $\beta$  is related to the amplitude of second harmonic, which is normalized by the square of the fundamental wave amplitude

$$\beta = \frac{8A_2}{A_1^2 k^2 x}. \quad (12)$$

In experiments, the quantity  $\hat{\beta}$  is measured:

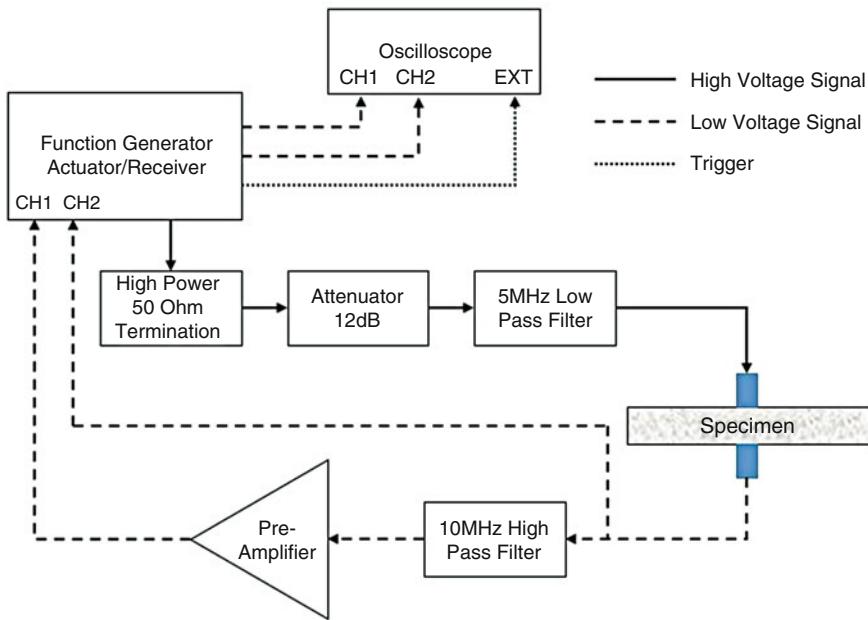
$$\hat{\beta} = \frac{A_2}{A_1^2} \propto \beta. \quad (13)$$

Thus, the material nonlinearity can be evaluated by detecting the fundamental and the second harmonic amplitudes of an ultrasonic test in a specimen.

An example for the nonlinear ultrasonic bulk wave NDT is provided in Fig. 2. A 5 MHz piezoelectric transducer (PZT) is employed to generate a signal with center frequency of 5 MHz. An attenuator and an amplifier are connected to the transmitting transducer and to the receiving transducer, respectively. The center frequency of the receiver is set at 10 MHz to obtain the corresponding second harmonic frequency component. Both transducers are carefully placed on each side of the specimen with holders designed to ensure uniform coupling conditions. As shown in Fig. 3, a Hanning window is imposed on the steady-state part of the signal, and signals are digitally processed by using the fast Fourier transform (FFT) to obtain amplitudes  $A_1$  at the fundamental frequency and  $A_2$  at the second harmonic frequency, respectively.

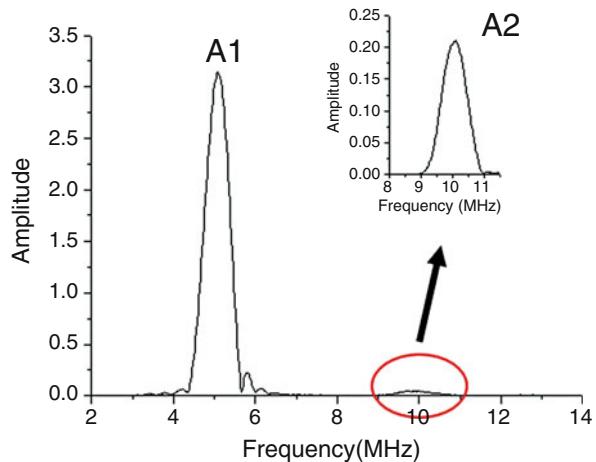
## Nonlinear Ultrasonic Guided Waves NDT

With the high sensitivity of the nonlinear ultrasonic approach and the advantages of guided wave techniques illustrated above, the nonlinear ultrasonic guided waves have drawn significant attentions for material characterization and micro-damage detection. Compared with bulk waves, second harmonic fields of guided waves are much more complex because of dispersion and their multimode nature. In general, the effect of second harmonic generation can be small and easily overlooked due to the dispersive nature of guided waves. Consequently, proper mode tuning with



**Fig. 2** Experimental setup for second harmonic generation of bulk waves

**Fig. 3** Typical measured primary and second harmonic wave amplitudes



physically based features is highly demanded to enhance the efficiency of nonlinear guided wave generation and reception.

Based on earlier investigations (Deng 1999; de Lima and Hamilton 2003; Bermes et al. 2007; Srivastava and di Scalea 2009; Li et al. 2012; Li and Cho 2016), it was found that the second harmonic amplitude grows linearly with the propagation distance at the internal resonant conditions of the second harmonic guided wave

mode and the primary guided wave mode. If the wave mode chosen satisfies these two conditions, the second harmonic amplitude will be cumulative. A series of double-frequency wave components will also be generated by the driving sources of nonlinearity. In practice, interest is focused on the second harmonic generation with the cumulative effect since the cumulative second harmonic plays a dominant role in the second harmonics field after a certain propagating distance.

The nonlinear parameter for symmetric Lamb wave modes can be represented by in-plate displacement on the surface as

$$\beta_{L.s} = \frac{A_2}{A_1^2} \frac{8}{k_l^2 x} \frac{\cosh^2(ph)}{\cosh(2ph)} \left( 1 - \frac{k^2 + q^2}{2k^2} \right), \quad (14)$$

where  $A_2$  and  $A_1$  are the in-plate displacement amplitude of the second harmonic mode and the primary wave mode with symmetric feature. The nonlinear parameter for anti-symmetric modes can also be derived with the same procedure as

$$\beta_{L.a} = \frac{A_2}{A_1^2} \frac{8}{k_l^2 x} \frac{\sinh^2(ph)}{\cosh(2ph)} \left( 1 - \frac{k^2 + q^2}{2k^2} \right). \quad (15)$$

Using the same method, the nonlinear parameters of the Lamb wave can be represented by the out-of-displacement on the surface as

$$\beta_{L.s} = \frac{A_2}{A_1^2} \frac{i8}{k_l^2 x} \frac{p}{k} \frac{\sinh^2(ph)}{\sinh(2ph)} \left( 1 - \frac{2k^2}{k^2 + q^2} \right), \quad (16)$$

$$\beta_{L.a} = \frac{A_2}{A_1^2} \frac{i8}{k_l^2 x} \frac{p}{k} \frac{\cosh^2(ph)}{\sinh(2ph)} \left( 1 - \frac{2k^2}{k^2 + q^2} \right), \quad (17)$$

where  $\beta_{L.s}$  and  $\beta_{L.a}$  are the nonlinear parameters for the symmetric and anti-symmetric Lamb modes, respectively. It is important to note that the cumulative second harmonic filed is symmetric even if the primary wave mode is antisymmetric.

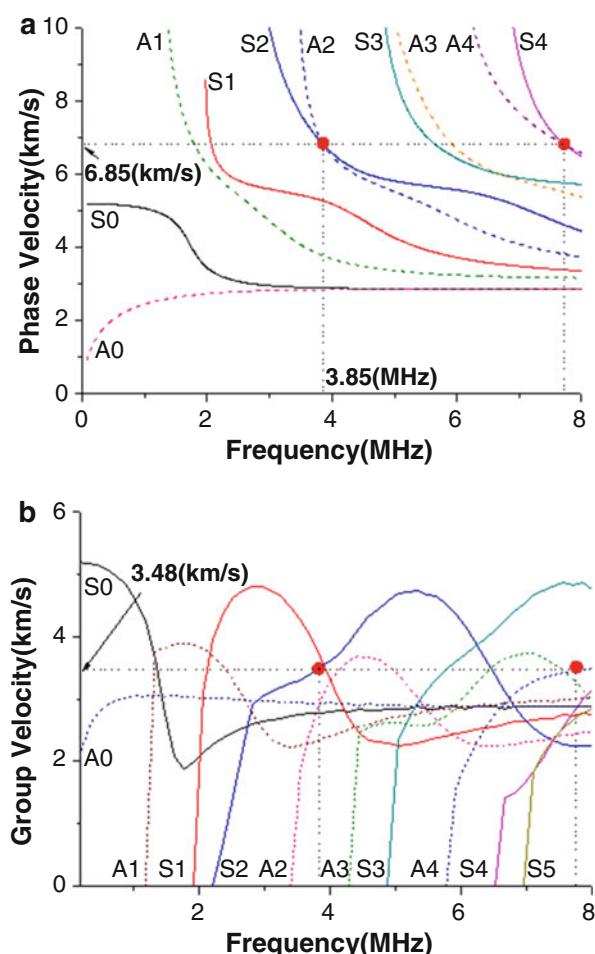
The formulas of the nonlinear parameter for the Lamb wave mode show that the acoustic nonlinear parameter for Lamb waves is a function of frequency, material properties, and geometric information of the waveguide. The nonlinear features of Lamb modes can be affected by the mode type, frequency of the incident signal, material properties, and the geometric information of the waveguide.

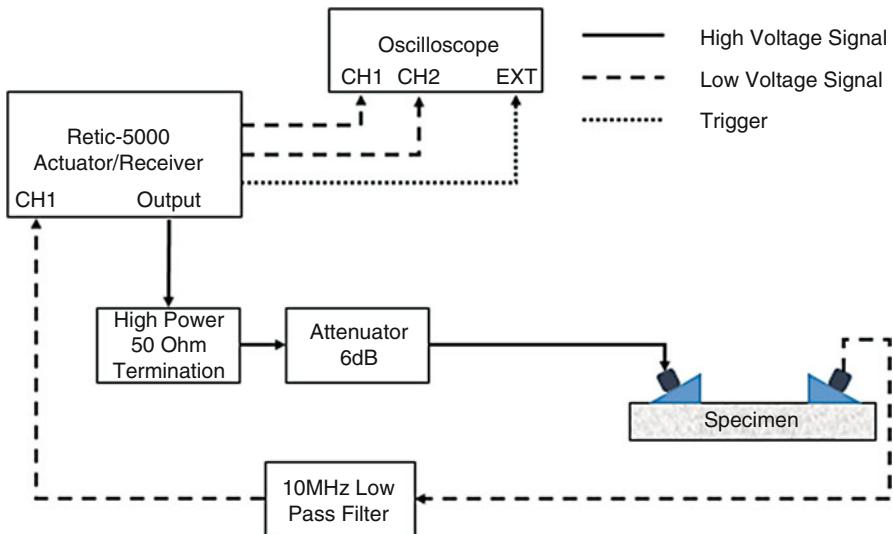
The second harmonic modes of guided waves display a cumulative effect under the conditions of phase matching and nonzero power transfer from the fundamental wave mode to the second harmonic wave mode. The cumulative effect of the second harmonic amplitude is of significant advantage for detection in experimental work to ensure measurement of the nonlinear effect with sufficient signal-to-noise ratio. In earlier investigation, phase-matching Lamb wave modes were chosen to evaluate material nonlinearities (Pruell et al. 2007; Deng and Pei 2007; Li and Cho 2014). The

concept of phase matching is based on the choice of the Lamb wave modes whose phase velocity equals that of the double frequency guided wave mode. The dispersion curves of Lamb wave propagation in the specimens used in this investigation were calculated numerically, as shown in Fig. 4.

Figure 5 shows the experimental setup used to generate and detect nonlinear guided waves. A high power termination was connected to the actuator to generate a tone burst signal of 20 cycles at a central frequency of 3.85 MHz. The generated sinusoidal signal then passes through the 6 dB attenuator, which was set to purify the signal to produce a high signal-to-noise ratio. A narrow-band contact piezoelectric transducer, whose nominal frequency is 2.25 MHz, was used to excite a longitudinal wave, and a receiver with central frequency of 5 MHz was set for the detection of the second harmonic wave centered at 5 MHz. The angle of the wedge is calculated using Snell's law. High vacuum

**Fig. 4** Phase velocity (a) and group velocity (b) dispersion curves for guided waves propagation in a stainless steel plate with 1.35 mm thickness





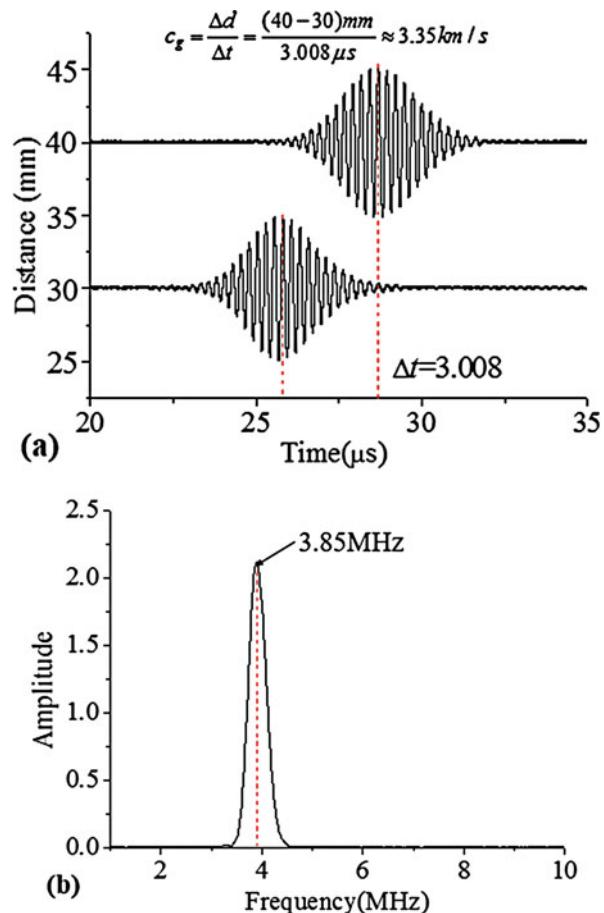
**Fig. 5** Experimental setup for nonlinear ultrasonic measurements

grease was used to acoustically couple the transducer and the wedge as well as the wedge and the specimen. The initial distance between the two wedges was set at 40 mm. The obtained time domain data recorded in the oscilloscope is processed using the fast Fourier transform after the signal passed through the 10 MHz low-pass filter.

Group velocity and frequency of the propagating signal are checked to identify the Lamb wave mode. Figure 6a shows the two typical waveforms of the received signal in time domain under different propagation distances in the undamaged specimen. The group velocity of the experimental signal is calculated as 3.32 km/s. The fundamental frequency of the signal is 3.85 MHz. Comparing the value with that shown in the group velocity dispersion curve in Fig. 6b, it can be shown that the propagating signal phase matches S2 Lamb wave mode.

A typical waveform of a received signal in the time domain resulting from the propagation in an undamaged specimen is shown in Fig. 7. Sixteen cycles of a sinusoidal signal (tone burst) are generated by a high power actuator and are modulated with a Hanning window. The received time-domain signal is processed in the frequency domain with the fast Fourier transform (FFT) to obtain its frequency spectrum; the existence of a second harmonic wave in the undamaged specimen demonstrates that there is nonlinearity in the specimen or measurement instruments. The plot in Fig. 8 shows the average data with error bars from the three measurement sets. The increase in the relative nonlinear parameter values with propagation distance shows the cumulative effect of the second harmonic generation.

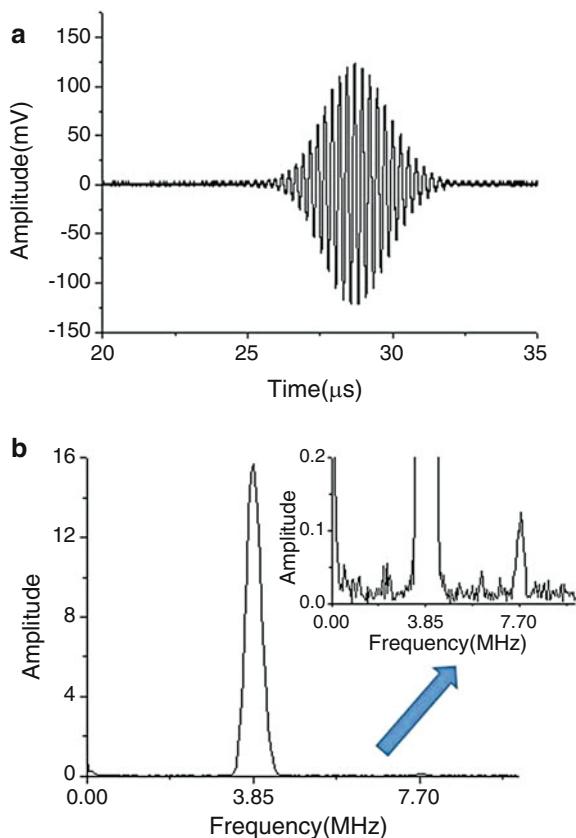
**Fig. 6** Group velocity (a) and frequency spectrum (b) of the experimental signal



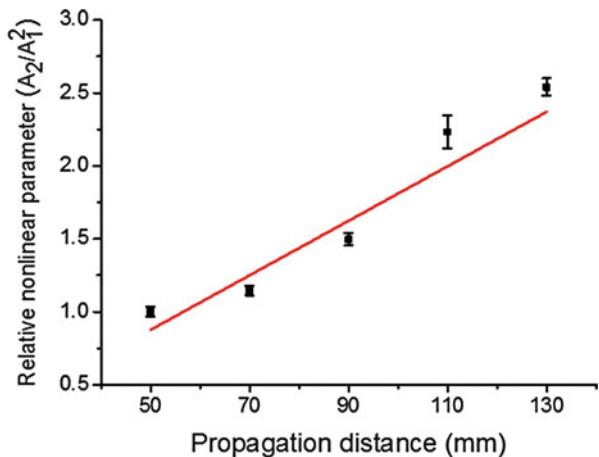
### Concluding Remarks

Various nonlinear phenomena accompany acoustic wave propagation in solid media with material nonlinearity. Even though the full mechanism of nonlinear response of waves caused by micro-damage is not yet well understood, the use of nonlinear acoustics can be a promising qualitative means for nondestructive testing, since acoustic nonlinearity is a much more sensitive indicator of micro-damage as compared to conventional linear ultrasonic features. The damage-induced material nonlinearity can be represented by the measure of acoustic nonlinearity with appropriate techniques, such as the second harmonic waves, frequency mixing response, as well as the subharmonic generation and nonlinear resonance frequency shift. It is also important to note that nonlinear features of various types of ultrasonic waves are also very different from each other; thus the guideline for proposing nonlinear ultrasonic technique for NDT should also be different.

**Fig. 7** Typical received signal: time domain signal and Fourier spectra of the fundamental and second harmonic signals



**Fig. 8** Relative nonlinear parameter  $\bar{\beta}$  versus propagation distances



## References

- Bentahar M, El Agra H, El Guerjouma R, Griffa M, Scalerandi M (2006) Hysteretic elasticity in damaged concrete: quantitative analysis of slow and fast dynamics. *Phys Rev B* 73(1):014116
- Bermes C, Kim JY, Qu J, Jacobs LJ (2007) Experimental characterization of material nonlinearity using Lamb waves. *Appl Phys Lett* 90(2):1–4
- Campos-Pozuelo C, Vanhille C, Gallego-Juárez JA (2006) Comparative study of the nonlinear behavior of fatigued and intact samples of metallic alloys. *IEEE Trans Ultrason Ferroelectr Freq Control* 53(1):175–184
- Cantrell JH, Yost WT (2001) Nonlinear ultrasonic characterization of fatigue microstructures. *Int J Fatigue* 23:S487–S490
- Chomas J, Dayton P, May D, Ferrara K (2002) Nondestructive subharmonic imaging. *IEEE Trans Ultrason Ferroelectr Freq Control* 49(7):883–893
- Croxford AJ, Wilcox PD, Drinkwater BW, Nagy PB (2009) The use of non-collinear mixing for nonlinear ultrasonic detection of plasticity and fatigue. *J Acoust Soc Am* 126:117–122
- de Lima WJN, Hamilton MF (2003) Finite-amplitude waves in isotropic elastic plates. *J Sound Vib* 265(4):819–839
- Demcenko A, Akkerman R, Nagy PB (2012) Non-collinear wave mixing for nonlinear ultrasonic detection of physical ageing in PVC. *NDT&E Int* 49(1):34–39
- Deng M (1999) Cumulative second-harmonic generation of Lamb-mode propagation in a solid plate. *J Appl Phys* 85(6):3051–3058
- Deng M, Pei J (2007) Assessment of accumulated fatigue damage in solid plates using nonlinear Lamb wave approach. *Appl Phys Lett* 90:121902
- Donskoy DM, Sutin AM (1998) Vibro-acoustic modulation nondestructive evaluation technique. *J Intell Mater Syst Struct* 9:765–771
- Donskoy D, Sutin A, Ekimov A (2001) Nonlinear acoustic interaction on contact interfaces and its use for nondestructive testing. *NDT & E Int* 34(4):231–238
- Eiras JN, Kundu T, Popovics J, Monzo J, Paya J (2014) Non-classical nonlinear feature extraction from standard resonance vibration data for damage detection. *J Acoust Soc Am – Express Lett* 135:EL82–EL87
- Favrie N, Lombard B, Payan C (2015) Fast and slow dynamics in a nonlinear elastic bar excited by longitudinal vibrations. *Wave Motion* 56:221–238
- Goldberg ZA (1956) On the propagation of plane waves of finite amplitude. *Sov Phys (Acoustics)* 2:346–352
- Guyer RA, Johnson PA (1999) Nonlinear mesoscopic elasticity: evidence for a new class of materials. *Phys Today* 52(4):30–36
- Guyer RA, McCall KR, Boitnott GN (1995) Hysteresis, discrete memory, and nonlinear wave propagation in rock. *Phys Rev Lett* 74:3491–3494
- Guyer RA, McCall KR, Van Den Abeele K (1998) Slow elastic dynamics in a resonant bar of rock. *Geophys Res Lett* 25:1585–1588
- Hamilton MF, Blackstock DT (1998) Nonlinear acoustics. Academic, London
- Herrmann J, Kim J, Jacobs LJ, Qu J, Littles JW, Savage M (2006a) Assessment of material damage in a nickel-base superalloy using nonlinear Rayleigh surface waves. *J Appl Phys* 99:124913
- Herrmann J, Kim J, Jacobs LJ, Qu J, Littles JW (2006b) Assessment of material damage in a nickel-based superalloy using nonlinear Rayleigh surface wave. *J Appl Phys* 99(12):1497–1488
- Hess P, Lomonosov AM, Mayer AP (2014) Laser based linear and nonlinear guided elastic waves at surfaces (2D) and wedges (1D). *Ultrasonics* 54:39–55
- Hikata A, Elbaum C (1966) Generation of ultrasonic second and third harmonics due to dislocations. *Phys Rev* 144:469–477
- Hilloul B, Abraham O, Loukili A, Durand O, Tournat V (2014) Small crack detection in cementitious materials using nonlinear coda wave modulation. *NDT & E Int* 68:98–104
- Hurley DC, Fortunko CM (1997) Determination of the nonlinear ultrasonic parameter using a Michelson interferometer. *Meas Sci Technol* 8:634–642

- Jhang KY (2000) Applications of nonlinear ultrasonics to the NDE of material degradation. *IEEE Trans Ultrason Ferroelectr Freq Control* 47:540–548
- Jhang KY, Kim KC (1999) Evaluation of material degradation using nonlinear acoustic effect. *Ultrasonics* 37:39–44
- Jia X, de Billy M (1992) Observation of the dispersion behavior of surface acoustic waves in a wedge waveguide by laser ultrasonics. *Appl Phys Lett* 61:2970–2972
- Johnson PA, Rasolofosaon PNJ (1996) Resonance and elastic nonlinear phenomena in rock. *J Geophys Res* 101(B5):553–564
- Johnson PA, Sutin A (2005) Slow dynamics and anomalous nonlinear fast dynamics in diverse solids. *J Acoust Soc Am* 117:124–130
- Kim J-Y, Baltazar A, Hu JW, Rokhlin SI (2006a) Hysteretic linear and nonlinear acoustic responses from pressed interfaces. *Int J Solids Struct* 43(21):6436–6452
- Kim JY, Qu J, Jacobs LJ, Littles JW, Savage MF (2006b) Acoustic nonlinearity parameter due to microplasticity. *J Nondestruct Eval* 25:28–36
- Klepa A, Staszewski WJ, Jenal RB, Szwedo M, Iwaniec J (2012) Nonlinear acoustics for fatigue crack detection – experimental investigations of vibro-acoustic wave modulations. *Struct Health Monit* 11:197–211
- Kober J, Prevorovsky Z (2014) Theoretical investigation of nonlinear ultrasonic wave modulation spectroscopy at crack interface. *NDT & E Int* 61:10–15
- Kuvshinov B, Smit T, Campman XH (2013) Nonlinear interaction of elastic waves in rocks. *Geophys J Int* 194:1920–1940
- Landau LD, Lifshitz EM (1970) Theory of elasticity. Oxford: Oxford University Press
- Li W, Cho Y (2014) Thermal fatigue damage assessment in an isotropic pipe using nonlinear ultrasonic guided waves. *Exp Mech* 54(8):1309–1318
- Li W, Cho Y (2016) Combination of nonlinear ultrasonics and guided wave tomography for imaging the micro-defects. *Ultrasonics* 65:87–95
- Li W, Cho Y, Achenbach JD (2012a) Detection of thermal fatigue in composites by second harmonic Lamb waves. *Smart Mater Struct* 21(8):085019
- Li W, Cho Y, Hyun S (2012b) Characteristics of ultrasonic nonlinearity by thermal fatigue. *Int J Precis Eng Manuf* 13(6):935–940
- Li W, Cho Y, Achenbach JD (2013) Assessment of heat treated Inconel X-750 alloy by nonlinear ultrasonics. *Exp Mech* 53(5):775–781
- Li W, Deng M, Xiang Y (2017) Review on the second harmonic generation of ultrasonic guided waves in solid media (I): theoretical analyses. *Chin Phys B* 26:114302
- Liu Y, Khajeh E, Lissenden CJ, Rose JL (2013) Interaction of torsional and longitudinal guided waves in weakly nonlinear circular cylinders. *J Acoust Soc Am* 133:2541–2553
- Liu P, Sohn H, Kundu T, Yang S (2014) Noncontact detection of fatigue cracks by laser nonlinear wave modulation spectroscopy (LNWMS). *NDT & E Int* 66:106–116
- Matlack KH, Kim J, Jacobs LJ, Qu J (2015) Review of second harmonic generation measurement techniques for material state determination in metals. *J Nondestruct Eval* 34:273
- Moreau A (1995) Detection of acoustic second harmonics in solids using a heterodyne laser interferometer. *J Acoust Soc Am* 98:2745
- Muller M, Sutin A, Guyer R, Talmant M, Laugier P, Johnson P (2005) Nonlinear resonant ultrasound spectroscopy (NRUS) applied to damage assessment in bone. *J Acoust Soc Am* 118(6):3946–3952
- Nagy PB (1998) Fatigue damage assessment by nonlinear ultrasonic materials characterization. *Ultrasonics* 36(1–5):375–381
- Nazarov VE, Radostin AV (2015) Nonlinear acoustic waves in micro-inhomogeneous solids. London: Wiley
- Nazarov VE, Radostin AV, Ostrovsky LA, Soustova IA (2003) Wave processes in media with hysteretic nonlinearity: part 2. *Acoust Phys* 49(4):444–448

- Ohara Y, Mihara T, Sasaki R, Ogata T, Yamamoto S, Kishimoto Y, Yamanaka K (2007) Imaging of closed crack using nonlinear response of elastic waves at subharmonic frequency. *Appl Phys Lett* 90:011902
- Ohara Y, Endo H, Mihara T, Yamanaka K (2009) Ultrasonic measurement of closed stress corrosion crack depth using subharmonic phased array. *Jpn J Appl Phys* 48:07GD01
- Padmore TC, Stegeman GI (1976) Surface-wave nonlinearities: nonlinear bulk wave generation by two oppositely directed collinear surface waves. *J Appl Phys* 47(4):1209–1228
- Pecorari C, Mendelsohn DA (2014) Forced nonlinear vibrations of a one-dimensional bar with arbitrary distributions of hysteretic damage. *J Nondestruct Eval* 33(2):239–251
- Pruell C, Kim JY, Qu J, Jacobs L (2007) Evaluation of plasticity driven material damage using Lamb waves. *Appl Phys Lett* 91:231911
- Qu J, Jacobs LJ, Nagy PB (2011) On the acoustic-radiation-induced strain and stress in elastic solids with quadratic nonlinearity (L). *J Acoust Soc Am* 129(6):3449–3452
- Read TA (1940) The internal friction of single metal crystals. *Phys Rev* 58:371–380
- Rischbieter F (1967) Measurement of the nonlinear sound response of aluminum with the aid of Rayleigh waves. *Acta Acoust United Acust* 18(2):109–112
- Rushchitsky JJ (2014) Nonlinear elastic waves in materials. London: Springer
- Scruby CB, Drain LE (1990) Laser ultrasonics: techniques and applications. Adam Hilger, Bristol
- Solodov IY (1998) Ultrasonics of non-linear contacts: propagation, reflection and NDE-applications. *Ultrasonics* 36:383–390
- Solodov IY, Korshak BA (2002) Instability, chaos, and “memory” in acoustic-wave-crack interaction. *Phys Rev Lett* 88:014303
- Srivastava AF, di Scalea L (2009) On the existence of antisymmetric or symmetric Lamb waves at nonlinear higher harmonics. *J Sound Vib* 323:932–943
- Stratoudaki T, Ellwood R, Sharples S, Clark M, Somekh MG (2011) Measurement of materials nonlinearity using surface acoustic wave parametric interaction and laser ultrasonics. *J Acoust Soc Am* 129:1721
- Sugawara A, Jinno K, Ohara Y, Yamanaka K (2015) Closed-crack imaging and scattering behavior analysis using confocal subharmonic phased array. *Jpn J Appl Phys* 54:07HC08
- Ten Cate JA, Shankl TJ (1996) Slow dynamics in the nonlinear elastic response of Berea sandstone. *Geophys Res Lett* 23:3019–3022
- Van Den Abeele KE-A, Johnson PA, Sutin A (2000a) Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage, part I: nonlinear wave modulation spectroscopy (NWMS). *Res Nondestruct Eval* 12:17–30
- Van Den Abeele KE, Carmeliet J, Ten Cate JA, Johnson PA (2000b) Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage, part II: single-mode nonlinear resonance acoustic spectroscopy. *Res Nondestruct Eval* 12:31–42
- Zaremba LK, Krasil'nikov VA, Shkol'nik IE (1989) Nonlinear acoustics in a problem of diagnosing the strength of solids. *Probl Prochnosti* 11:86–92



# Local Acoustic Resonance Spectroscopy

9

Christian U. Grosse, Anne Jüngert, and Philipp Jatzlau

## Contents

Introduction .....	272
LARS Compared to Other NDT Techniques .....	272
Technical Prerequisites and Application Concept .....	274
Basic Principles and Theoretical Background .....	274
Body/Body Collisions Described Using Hertz Impact Theory .....	275
Theoretical Background to Investigate Materials .....	276
Impactor Stiffness and Influence of Impact Force .....	277
Instrumentation and Setup .....	279
Emitter/Impactor .....	279
Receiver/Microphone .....	280
Recording Equipment and Automation .....	281
Signal Processing .....	282
Applications .....	284
Applications for Wind Turbine Blades .....	284
Investigation of Smaller Impact Specimens Made of Glass Fiber-Reinforced Polymers .....	287
Applications to Test Carbon Fiber-Reinforced Polymers .....	287
Summary and Outlook .....	291
References .....	293

---

C. U. Grosse (✉)

Center for Building Materials, Chair of Non-destructive Testing, Technical University of Munich, München, Germany

e-mail: [grosse@tum.de](mailto:grosse@tum.de)

A. Jüngert

University of Stuttgart, Stuttgart, Germany

e-mail: [anne.juengert@mpa.uni-stuttgart.de](mailto:anne.juengert@mpa.uni-stuttgart.de)

P. Jatzlau

Center for Building Materials, Non-destructive Testing, Technical University of Munich, München, Germany

e-mail: [philipp.jatzlau@tum.de](mailto:philipp.jatzlau@tum.de)

**Abstract**

The local acoustic resonance spectroscopy (LARS) is a novel nondestructive testing (NDT) technique that is based on the well-known coin-tapping test. This rather old test utilizes the effect that a defect (such as a large void, crack, or delamination), that is invisible from the surface, causes a change of acoustic waves emitted from the structure as a person taps on it – e.g., using a coin or a small hammer. This procedure has obvious drawbacks as it depends strongly on the skills – tapping correctly and listening carefully to the sound emitted – of an experienced person who needs to apply this simple technique by hand being close to the structure. Proper interpretation of the emitted sound is often difficult – even for experienced investigators. However, the benefits of easy implementation, relatively fast surveying, and cost-efficiency are evident as well. This was the motivation for the instrumentation of this technique, where all steps are automated, including a contact-free recording microphone and a robot-based instrumented excitation of the component under test. The technique can be locally applied to structures constructed from fiber-reinforced polymers and uses the acoustic waves that are analyzed in the frequency domain according to resonance peak content. For industrial applications, LARS can be used as a rapid but reliable NDT technique to obtain an initial overview about larger defects in a structure prior to implementing more detailed investigations, thus raising the escalation level of quality control.

---

## Introduction

Mechanical impacts can cause invisible damage inside a material, a problem in particular for structures constructed from fiber-reinforced polymers. Such defects, triggered, for example, by compression after impact (CAI), lead to partial delamination, which may have significant influence on the mechanical properties of such materials (Abbate 2005). The damage caused by a so-called tool drop scenario is relevant in the automotive and aeronautical sectors (Grosse et al. 2016) and also for rotor blades of wind turbines. It requires techniques that can detect such defects in a reliable as well as rapid and cost-efficient way – local acoustic resonance spectroscopy (LARS) as a more mechanically automated coin-tapping test can be used in this scenario. The development of a novel test procedure such as LARS may require a more detailed comparison to other existing techniques and a deeper evaluation of benefits and challenges. The comparison and evaluation is outlined in the following subsections along with a more general description of possible applications before this is addressed together with the basic principles in the main sections that follow.

### LARS Compared to Other NDT Techniques

The dominant nondestructive testing (NDT) technique today using acoustic waves is ultrasound. The waves are evaluated in the frequency spectrum above the audio range ( $>20$  kHz). Ultrasonic waves can be used in transmission or reflection modes and are characterized by an emitter and receiver (mostly using the piezoelectric

effect) that are used in separated housings or are combined in one housing. Ultrasonic techniques are applied at a single position (A-scan), at multiple positions along a line (B-scan), or in several parallel lines (C-scan). Advanced test setups utilize several sensors in array configurations or in one housing to apply measuring and data processing techniques such as phased array, synthetic aperture focusing techniques or total focusing methods. The emitter and receiver are usually coupled to the surface of a specimen using a coupling agent, which is a significant difference to LARS. Another difference is the sound excitation that is done with tapping the surface. A broad range of frequencies is radiated into the material as concentric elastic waves. The generation of waves in a broad frequency range using piezoelectric sensors is a bit more complicated.

Vibrational testing is another NDT method that is used in quality control and in the evaluation of structures. Sensors (coupled to the surface or contact free) can determine the eigenvibrations of a structure excited by ambient conditions (wind, traffic, etc.) or artificial means (shaker, pulse hammer, etc.). Structures are – other than LARS – excited as a whole so that they vibrate according to the eigenvalues of the entire structure. Deviations from formerly calculated resonance frequencies are an indicator of defects.

LARS is most similar to the impact-echo (IE) technique that was established in the 1980s (Carino et al. 1986; Sansalone and Streett 1997). In IE the impulse is generated by a ball drop and received at the location of excitation by a conventional broadband piezo-sensor. A structure is investigated locally using changes of the local maximum frequency of a wave that can theoretically be explained (in cases of platelike structures) to be the  $S_1$  lamb mode with zero group velocity (Gibson and Popovics 2005). Applications are not restricted to the ultrasound spectrum but can be done at lower frequencies as well. Most commercially available IE devices use steel balls as an impulse source and have a sensor coupled to the surface near the point of impact where the pulse is generated (i.e., at a distance of a few centimeter) around the sensor. LARS instead uses microphones to record the response of the structure to the impulse. The IE technique was further developed to use microphones as well (Zhu and Popovics 2007) or to use an array of microphones or microelectromechanical systems (MEMS)-based microphones (Groschup and Grosse 2015; Ham and Popovics 2015). Most applications of IE that are found in the literature are limited to construction materials, namely, concrete and concrete slabs, with typical thickness of slabs in the range from approximately 10 cm up to half a meter. LARS techniques as referred to in this chapter are applied to much thinner plates and structures mostly out of fiber-reinforced polymers.

There are already several commercially available devices that facilitate – at least partially – the routine use of LARS. One is the so-called *Woodpecker* designed by Mitsui company (Mitsuhashi et al. 1989). Another is the *Bondmaster* (distributed by Olympus company (2010)), in which probes are used that are based on a so-called mechanical impedance analysis (MIA) technique (Fig. 1, left). MIA is a low-frequency nonresonant inspection method, which uses a single-tipped, dual-element probe. The dual-element probe (generating audible sound waves (1 to 10 kHz) into the test specimen) consists of a driver element that is coupled to the test piece through a plastic cone and wear shoe. The probe assembly is shown in Fig. 1, right. The company offers ten different MIA probes. The main difference between them is the tip diameter, which varies between 6.35 mm and 12.7 mm. The



**Fig. 1** Bondmaster, to which different probes can be connected (left), among them the MIA probe (detail: right)

system contains a spring-loaded probe shoe, does not require a couplant, and has a small contact area so it can be used on irregular or curved surfaces.

## Technical Prerequisites and Application Concept

LARS applications are per definition restricted to a local excitation of a larger structure using an instrumented pulse hammer. The forces should be limited to a level so as not to damage the surface. Receivers that are most commonly used are microphones with a wide frequency range in the audio spectrum. The technique is developed from the well-known coin-tapping test. It is optimized for rapid first inspection of structures with deviations of the audio response in a scanning line used as the basis to select areas for further, more detailed, investigations of suspicious spots. This makes LARS an ideal first method in an escalation approach where inspection techniques that are more detailed but laborious can follow to more precisely determine the defect geometry and type. Defect types that can be detected are delamination, larger voids, and cracks parallel to the surface. Some examples of applications to detect such defects in fiber-reinforced polymers are given in the following sections.

---

## Basic Principles and Theoretical Background

Applying LARS, the deflection of the surface after impact (using a ball or a hammer) is observed. This limits its applicability since soft materials cannot be investigated, if the deflection time is long compared to the impulse or is even of the plastic type. Therefore, only materials with at least an intermediate stiffness are considered. The

relevant parameter is the contact time between the impactor and the surface. The lower the stiffness of the material, the longer is the contact time. A defect at a certain depth reduces the stiffness leading to a higher contact time.

## Body/Body Collisions Described Using Hertz Impact Theory

The excitation pulse on a surface can be considered as the interaction of two bodies. Such a phenomenon can be described using the Hertz impact theory first developed by Heinrich Hertz (1881), who described applications of his theory concerning the interaction of two elastic steel spheres. The basic aspects can be found in later publications as well (Landau and Lifschitz 1989). The deflection  $\delta$  of such bodies depends on the contact stiffness  $k_c$  and the normalized force  $F$  as follows:

$$\frac{1}{k_c} = \frac{d\delta}{dF} \quad (1)$$

The deflection itself is governed by the material properties of the two bodies such as Young's modulus  $E$  and Poisson's ratio  $v$  and is also influenced by geometric parameters. The contact stiffness can, according to Cawley and Adams (1988) and Jüngert et al. (2013), be derived for a semi-infinite solid structure impacted by a sphere with radius  $r$ :

$$\delta = \left( \frac{9F^2}{16rE^{*2}} \right)^{1/3} \quad (2)$$

Following Shi and Polycarpou (2005),  $E^*$  is related to the sum of the elastic moduli and Poisson's ratios of the two bodies using the relation:

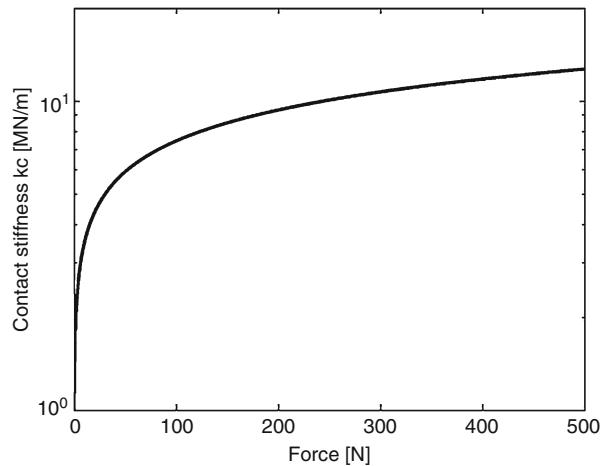
$$\frac{1}{E^*} = \frac{1 - v_1^2}{E_1} + \frac{1 - v_2^2}{E_2}. \quad (3)$$

Differentiation of the deflection with respect to the force  $F$  results in the reciprocal contact stiffness between the sphere and the semi-finite solid:

$$k_c = \sqrt[3]{\frac{3}{2} \left( \frac{16rE^{*2}}{9} \right) \cdot F} \quad (4)$$

Typically, the deflection follows a nonlinear relationship with the applied forces in such a way that higher impact forces result in a higher contact stiffness. Therefore, the applied forces using LARS methods should be more or less constant. For the simple case of a steel sphere with  $\mathcal{O} = 8$  mm,  $E_1 = 210$  kN/mm<sup>2</sup>, and Poisson's ratio  $v_1 = 0.29$  impacting a semi-infinite solid made of glass fiber-reinforced polymers ( $E_2 = 13$  kN/mm<sup>2</sup> and Poisson's ratio  $v_2 = 0.25$ ), the contact stiffness is described by the curve in Fig. 2.

**Fig. 2** Contact stiffness of an interaction of a steel sphere with a GFRC plate depending on the excitation force (Jüngert 2010)



As depicted in Fig. 2, the relation follows a third root law leading to a strong dependence of the contact stiffness on the excitation forces. These relations can be transferred from a sphere/plate interaction (McLaskey and Glaser 2010) to the interaction of an impulse hammer with a plate. The influence of the impact force value on the contact stiffness measurements is evaluated in section “[Impactor Stiffness and Influence of Impact Force](#).”

## Theoretical Background to Investigate Materials

Using LARS in materials testing applications, the contact stiffness between an impact hammer and the specimen is investigated through a change of the resulting sound waves. The effective stiffness is, according to Cawley and Adams (1988), a combination of the stiffness  $k_c$  in the undisturbed area and the stiffness  $k_d$  of the defective material regions. For a concentric defect with an aperture  $d$  at a depth  $h$ ,  $k_d$  can be derived from:

$$k_d = \frac{64\pi D}{d^2}, \quad (5)$$

where  $D$  can be calculated from Young’s modulus and Poisson’s ratio of the material according to:

$$D = \frac{Eh^3}{12(1 - \nu^2)}. \quad (6)$$

The defect stiffness is therefore proportional to the third power of the defect depth and inversely proportional to the diameter of the defect squared. This can be illustrated by an example: A defect with the diameter of 5 mm located at a depth

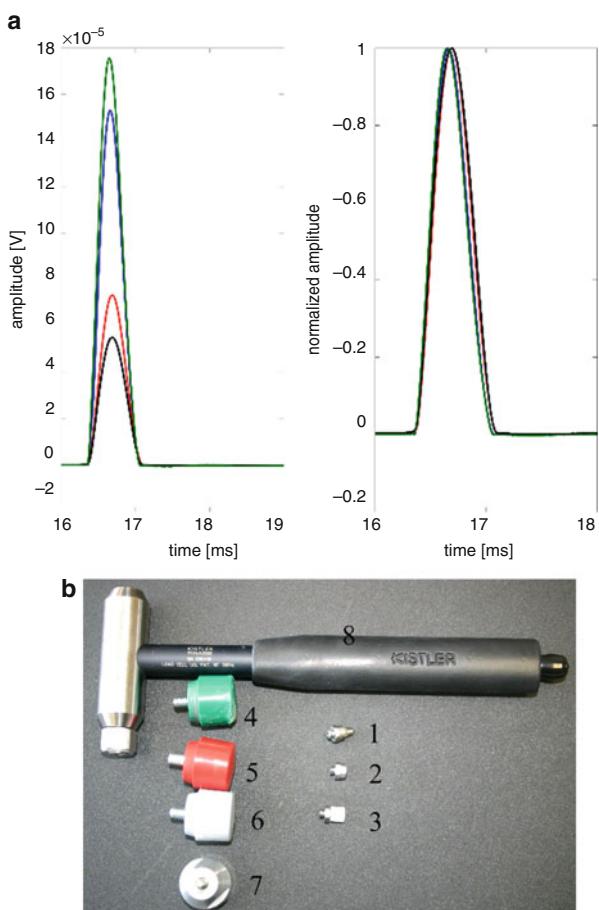
of 1 mm results in the same change of effective stiffness as a defect of size 26 mm in a depth of 3 mm. This demonstrates a much higher sensitivity of the technique to shallow defects. As mentioned above, a reduction in effective stiffness causes a longer contact time between the impactor and plate. Using an instrumented impulse hammer instead of a sphere, the force-time signal of the hammer is stretched, resulting in a smaller frequency spectrum and a different acoustic response.

## Impactor Stiffness and Influence of Impact Force

As described in section “[Body/Body Collisions Described Using Hertz Impact Theory](#),” a more or less reproducible impact source with controlled energy excitation is required. The use of spheres (made of steel or other materials such as rubin) is possible since three relevant parameters (frequency spectrum, stiffness of the sphere, and released energy) can be varied. The stimulated frequency spectrum is influenced by the diameter of the sphere, while the strength and stiffness can be modified using sphere materials with different hardness. Changing the size and therefore the mass as well as the drop height influences the released energy. However, repeatability and handling using spheres as sources of the impact are not as good compared to instrumented impulse hammers as they are used in modal analysis. The excitation force can be manually varied in an easy way, since it is recorded as a time function using a load cell that is built into the head of the hammer. The stiffness can be modified using hammer tips with different hardness. A softer tip results in a lower contact stiffness. As described in Jüngert (2010), the hammer impact produces under ideal conditions a bell-shaped curve where the energy of the hammer impact is given as the area beneath the curve. The impact on a stiff surface (high contact stiffness) will lead to a high and narrow bell curve.

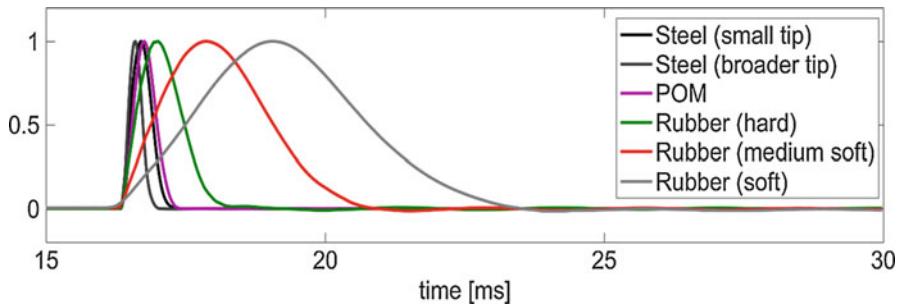
The impact on a soft material (low contact stiffness) will produce a broader curve. Based on similar excitation forces, the area beneath the curves should be equal. Thus, the width of the bell curve depends on the contact stiffness. Flaws such as delaminations or voids decrease the contact stiffness and therefore lead to broader force signals. To demonstrate these effects, an instrumented impulse hammer was used to manually tap on a flawless glass fiber-reinforced polymer (GFRP) plate. Different hammer tips were used – small, hard tips made of steel and plastic and larger tips made from elastomers. In Fig. 3a, the excitation signals for different forces using a polyoxymethylene (POM) tip (trade name Delrin™) are shown. The leftmost graph displays the recorded forces over time. The differences in the excitation force are visible in the different amplitudes of the signals. The smallest amplitude is more than three times smaller than the highest amplitude. Eq. 2 predicts a change in the contact stiffness. This change should lead to a change in the contact time between the hammer and GFRP plate. Therefore, it is necessary to look at the widths of the different excitation signals. The right graph of Fig. 3a shows the normalized excitation signals for a POM tip. It can be seen that differences in width are visible but small. Looking closer into the data (Jüngert et al. 2013), it can be seen that the half-widths of the curves vary by around 50 µs from the lightest to the strongest tap.

**Fig. 3** (a) Signals recorded by the load cell using a POM hammer tip and different excitation forces. (b) Impulse hammer with tips constructed from different materials (#3 is POM)

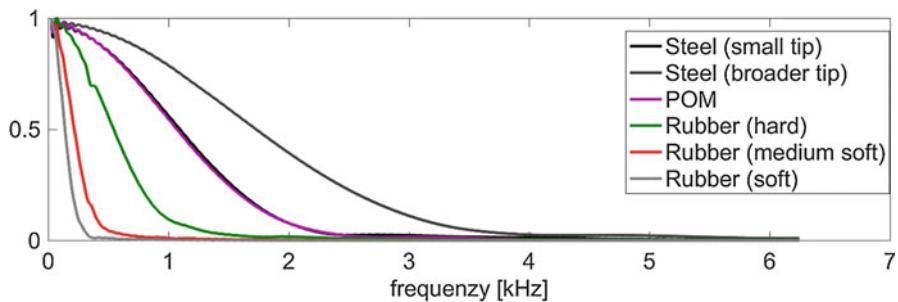


This difference is at least one order of magnitude smaller than differences caused by damages within GFRP.

How does the force-time graph (and the emitted frequency spectrum) change with different tip material? This was investigated using different tip materials like steel (with a small and a broader tip), POM (Delrin), and rubber (using a soft, medium, or hard elastomere) (Fig. 3, right). Figure 4 shows the force/time curves of the hammer impacts. The bell-shaped curves become, as expected, broader and broader the softer the material of the tip is (however, the small steel tip is not representative here, since it produced a nonelastic impact). This has also a direct effect on the emitted frequencies as shown in Fig. 5. As the hardness of the tip increases, higher frequencies are emitted and the frequency band becomes broader. However, the tip of the hammer should in case of high hardness not be too small to avoid any damage to the structure (as is the case for the small steel tip).



**Fig. 4** Contact time of hammer tip with a GFRC plate using hammer tips of different materials (see Fig. 5)



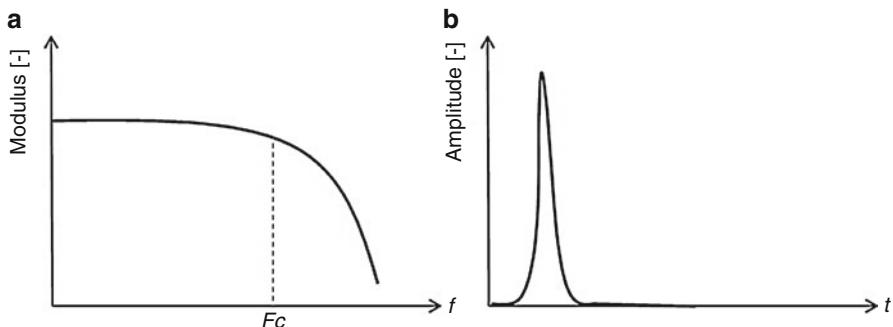
**Fig. 5** Normalized amplitude spectra of excitations using different hammer tip materials

## Instrumentation and Setup

The equipment typically used for LARS is a combination of devices and sensors that can also be used for vibrational tests and for impact-echo or ultrasound applications. In practice, to setup a LARS measurement, the most suitable emitter (usually a hammer) or sensor type depends strongly on the defect size and type and of the stiffness and geometry of the material to be tested. The different parts of the setup are described in the following sections.

### Emitter/Impactor

It is essential to apply a pulse that is transient and produces waves that contain a broad frequency range as demonstrated in Fig. 6. The shorter the pulse (Fig. 6, right), the higher is the corner frequency  $F_c$  (Fig. 6, left). The optimum would be to generate a pulse similar to the mathematical Dirac's delta function with a linear distribution of frequency amplitudes. In the real world, such a pulse cannot be generated because significant energy is required to cause locally a vibration of the structure. A compromise



**Fig. 6** Typical characteristic frequency content (left) of a hammer pulse and the source time function

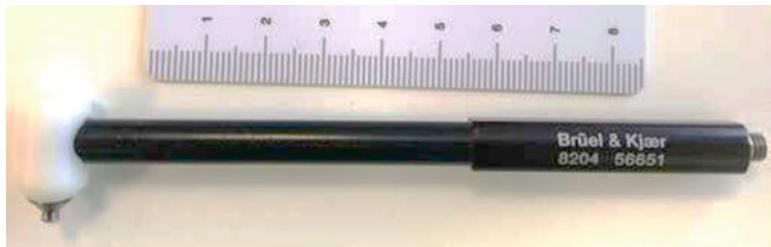
is needed. Using a piezo-electric transducer is not suitable since even broadband transducers will have a frequency transfer function showing resonances. Moreover, they need to be coupled to the surface to transmit the pulse into the structure which is contradictory to the concept of tapping and prohibits a rapid test. As described in the first two sections of this chapter, spheres are one way to generate such a pulse since the potential energy can be controlled via measurements of the drop height. However, reproducibility for some technical surfaces can be low due to surface roughness.

Another good solution is instrumented hammers as they are used in vibrational or modal analysis testing. Such a hammer – as shown in Fig. 2, right – is easy to handle. The energy of the pulse can be adjusted according to the boundary conditions, the material to be tested, and the hardness of the surface. The force function and – more important – the contact time can be recorded through the load sensor that is typically part of the hammer head. The data can be normalized using the force values with respect to the pulse energy. Double rebound of the hammer (tapping twice in the recording window) can also be suppressed as well as forces that lead to critical (i.e., plastic) deformations of the surface by reducing the mass of the hammer.

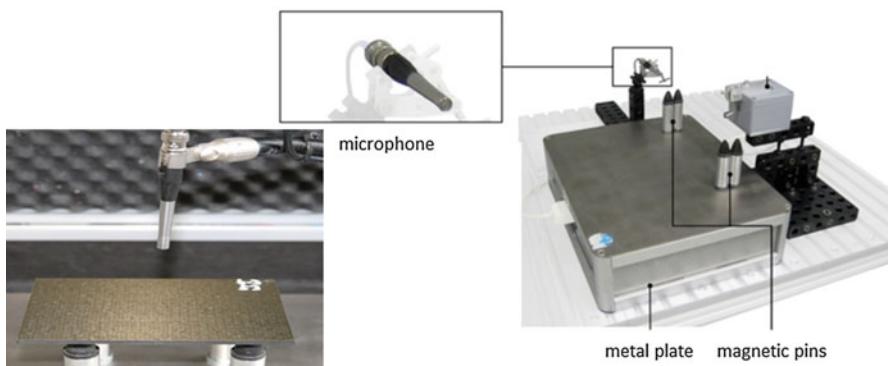
Since the contact time is directly related to the emitted frequency spectrum, the hammer and in particular the hammer tip can be chosen according to the material properties. In the case of structures with high hardness and small thickness, a modal analysis hammer is chosen that exhibits low mass and a steel tip for proper contact stiffness. As described in a later section, structures constructed of carbon fiber-reinforced polymers can be tested with such a smaller hammer (Fig. 7). For structures such as rotor blades constructed of glass fiber-reinforced polymers, a medium-size hammer as described in section “[Impactor Stiffness and Influence of Impact Force](#)” is most suitable.

## Receiver/Microphone

To develop fast measurements for large structures, contact-free recordings of the pulse response are required. An airborne recording of the acoustical wave emitted at the location of the pulse can be realized by conventional microphones. Typically



**Fig. 7** Small instrumented modal analysis hammer



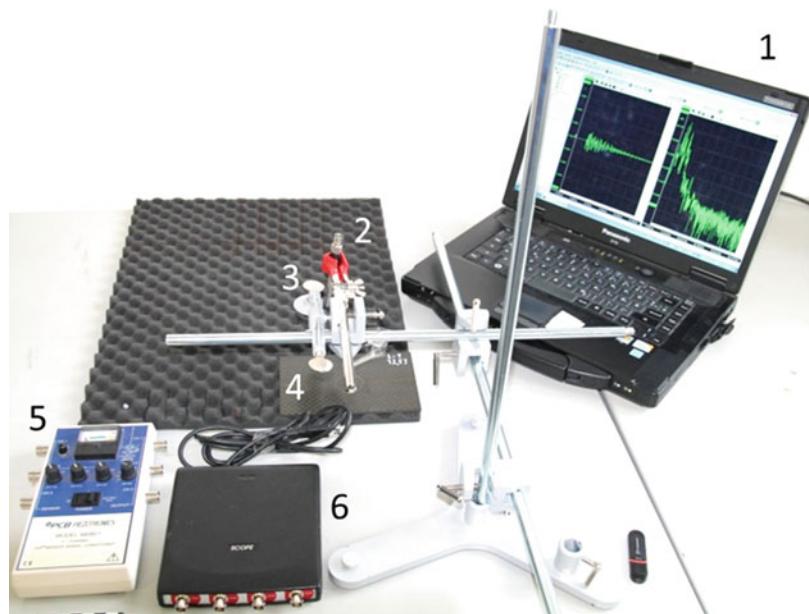
**Fig. 8** Setup for LARS using a 1/2" microphone and sample holders (Müller 2015)

a 1/2" prepolarized free-field condenser microphone is appropriate. Such microphones (see section "[Applications](#)") can record frequencies between 3.5 Hz and 20 kHz with a linearity better than  $+/- 2$  dB. An example of such a setup is shown in Fig. 8.

An alternative recording system would be a laser Doppler vibrometer measuring the out-of-plane vibrations excited by the hammer or the ball drop. The laser spot is focused close to the impact point and can be moved rapidly to the next measuring point if necessary. However, such a system is less flexible and certainly more expensive than a microphone-based one.

## Recording Equipment and Automation

LARS applications as developed during the last few years can be subdivided into two setups. One uses an instrumented modal analysis hammer and the other uses an automated ball drops. As for the latter, an apparatus with a ball holder and a magnet can be used (Fig. 9). A microphone (see section "[Receiver/Microphone](#)") is attached to the ball holder and its signals are converted by an A/D converter for automated recording and data processing.



**Fig. 9** Ball-drop setup for LARS measurements with (1) laptop with multichannel software, (2) microphone, (3) ball-drop device, (4) specimen, (5) DC source, and (6) digital oscilloscope TiePie HS4 (Müller 2015)

The setup using a miniature modal analysis hammer is depicted in Fig. 10. The hammer is released and mechanically moved across the specimen using a robot.

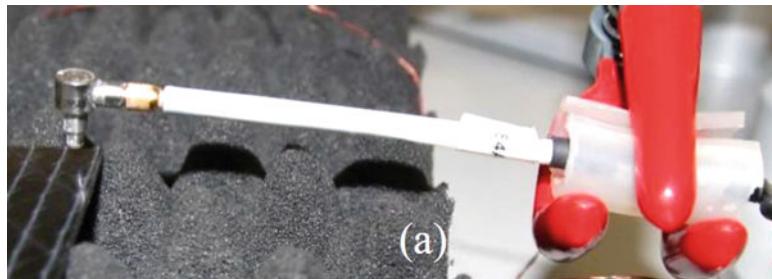
This setup can be modified using an automated pre-tensioning loading mechanism to realize a hammer pulse that is better reproducible with more or less constant pulse energy. Figure 11 shows the individual parts of such a system as developed by Narr (2017). The hammer (1) is “loaded” through a rubber band mechanism (3) and guided onto the specimen (7) using a hinge (5).

For the digitization an A/D converter with lower digitization rate can be used. If the maximum resonance frequency is in the range from 10 to 20 kHz, the digitizer should have a ten times higher sampling rate to match Shannon’s sampling theorem and to avoid the aliasing effect. In addition, an analogue low-pass filter would be beneficial.

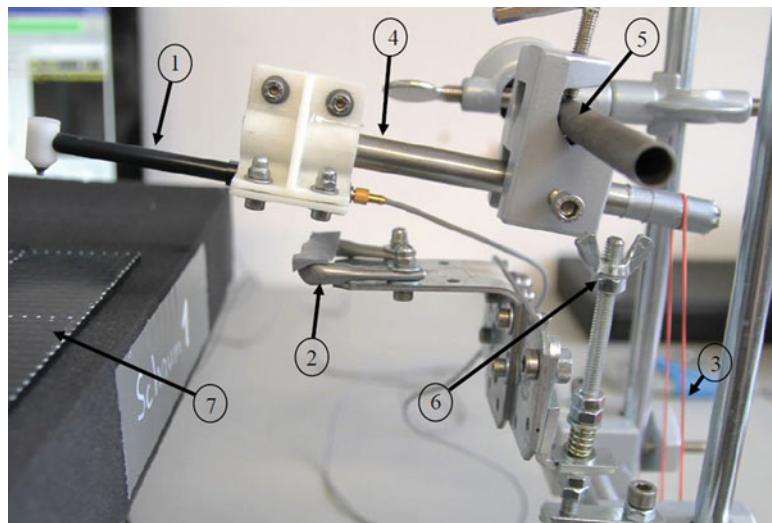
## Signal Processing

As described in the second section of this chapter, two parameters, both linked to the contact stiffness between test material and hammer tip, are analyzed: the force signal recorded at the hammer tip and the excited sound.

The force signals recorded in the hammer tip using a dynamic force sensor are ideally shaped like Gauss curves. The Fourier transform of a Gauss curve leads again to a Gauss curve; therefore, the analysis could be done either in the frequency



**Fig. 10** Hammer-based LARS setup using a miniature impact hammer (a) (Narr 2017)

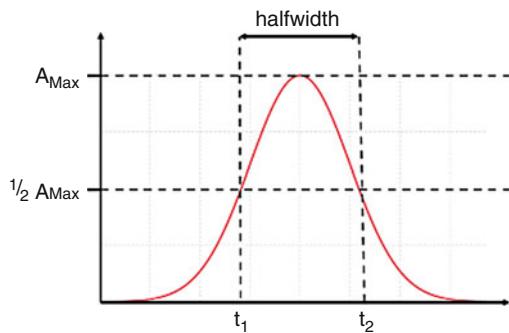


**Fig. 11** Automatic impact hammer device – details are described in the text and in Narr (2017)

domain or in the time domain. As the data are recorded in the time domain, the analysis will be done in the time domain as well. The simplest way to visualize the changes of the contact stiffness in the excitation signals is to determine the half-widths of the excitation signals. The tap on a stiff material generates a narrow curve with high amplitude; the tap on a soft material generates a broad curve with lower amplitude. The half-widths of the curves are defined as the width values measured at half of the maximum peak values as shown in Fig. 12. The half-width values are calculated from the measured data and plotted using the measurement grid in a color-coded or grayscale two-dimensional image for easier interpretation.

The recorded sound signals are analyzed in the frequency domain. Common coin-tapping tests are based on the effect that internal damage and cavities locally modify the contact stiffness and therefore alter the received acoustic signal. Audible changes in sound in general can be seen as changes in the frequency spectra of the recorded sound signals. The recorded sound signals are transferred into the frequency domain

**Fig. 12** Determination of half-width values



using a fast Fourier transform determining the amplitude spectra for each tap. In relation to the two-dimensional (2D) measurement field, a data cube is generated, which contains amplitude and frequency information for each measurement point on the 2D grid. Depending on the measurement data, different parameters can be chosen from the data cube, to generate color-coded 2D images, e.g., for different frequency ranges. If LARS is applied along a scanning line, it is useful to display the amplitude spectra as color-coded B-scans along the measurement grid as shown in Fig. 14.

## Applications

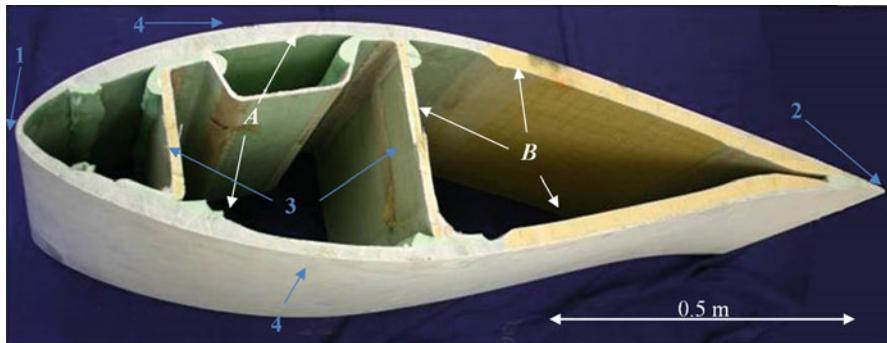
### Applications for Wind Turbine Blades

Wind turbine blades are highly stressed components and therefore need to be inspected regularly. However, current inspection techniques are limited to visual inspections and manual tapping tests. These manual tapping tests can be replaced by LARS.

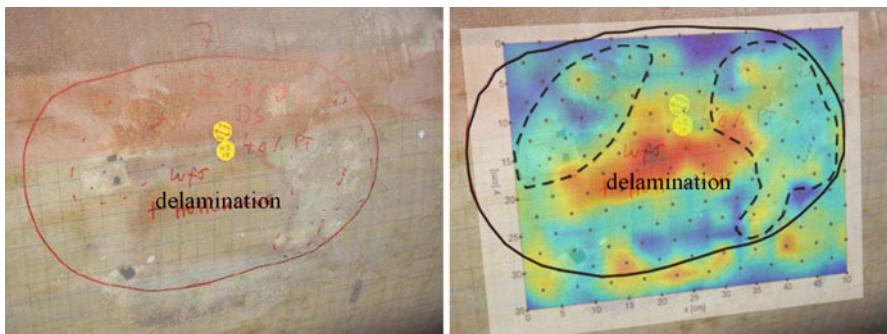
Wind turbine blades are made of (mainly glass) fiber-reinforced plastics combined with lightweight materials such as balsa wood or plastic foam. The combination of high sound scattering in the thick glass fiber laminates and the high sound damping in the lightweight parts of the wind turbine blades presents challenges for nondestructive testing.

A typical cross section of a wind turbine blade is shown in Fig. 13. LARS can be applied to both sandwich materials and to GFRP laminates. As described in section “[Basic Principles and Theoretical Background](#),” LARS is more sensitive to flaws near the surface.

Typical damage that occurs in wind turbines during operation or during manufacturing differs from typical damage in aerospace structures. Crucial damages are delaminations and cavities within the fiber layers, delamination of different materials, flaws in the bond lines, fiber undulations, and cracks in all materials. LARS is suitable to detect delaminations, cavities, and cracks parallel to the surface. It was not yet applied to fiber undulations as far as we know.



**Fig. 13** Cross section of a wind turbine blade; the upper side is the intake side; the lower side is the pressure face; the leading (1) and the trailing edge (2) can be seen; the spars (3) are glued to the belts (4); shaping areas are constructed of GFRP (**a**); for weight reduction some regions (**b**) are built as sandwich construction



**Fig. 14** Left, marked delamination in wind turbine blade; right, results of the analysis of the force signals recorded at a damaged area of a wind turbine blade

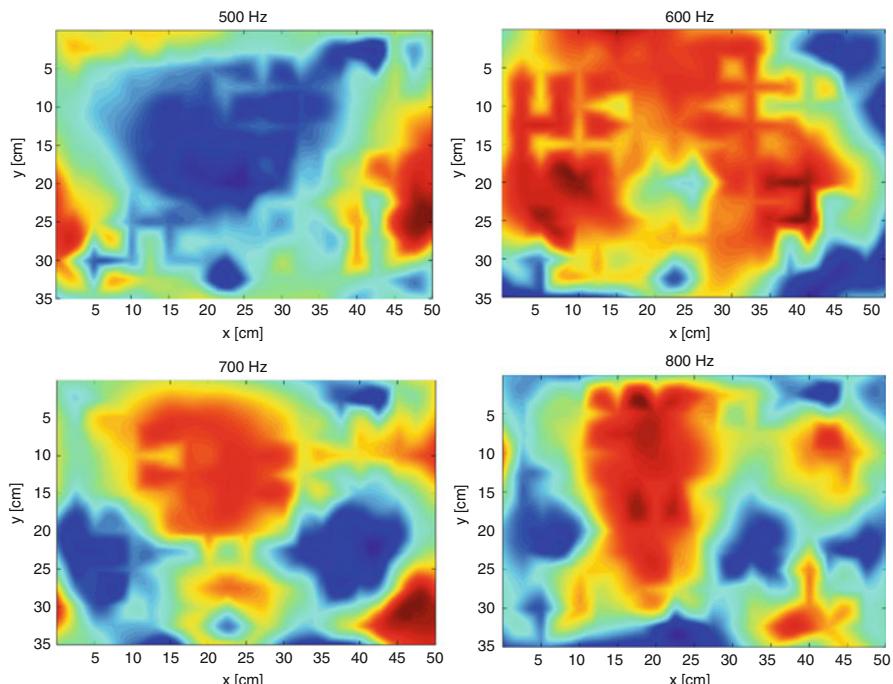
A typical field of application of LARS is for wind turbine blades, containing areas with visible or audible material changes. Figure 14, left, shows one of these areas. In the middle of the encircled area, a delamination was marked that was detectable using the conventional coin-tapping test. Within the encircled area, other regions are marked with dashed lines. Here near-surface material changes that are visible are caused by humidity.

The right part of Fig. 14 shows the damaged section of the blade and the results of analysis of the excitation forces as an overlay plot. Areas shown in red represent the areas with elongated contact times between hammer and impact. This is caused by a weakening of the material. It can be seen that the red area corresponds well to the earlier detected deterioration in the middle of the graph. It is assumed that the delaminated area is extended into the areas encircled with dashed lines. However, these defects might be camouflaged by the deterioration that is visible at the surface as the material weakens. These material changes are likely caused by water ingress

leading to embrittlement of the material. The brittle material causes shorter contact times. As the embrittlement moves closer to the surface, the increase in contact times due to the deeper delamination is no longer visible.

The analysis of the sound signals is usually done in the frequency domain. For better comparison of the data, normalization with respect to the actual impact force is required. After the normalization, the fast Fourier transform is applied to the sound data, and the amplitude spectra are calculated using absolute values of the complex spectra.

Figure 15 shows the distribution of spectral amplitudes for the observed blade area using four different primary frequencies. Again, the delamination is assumed to be in the center of the tested area. In the frequency range between 400 Hz and 1000 Hz, significant changes in the amplitude distribution are noticeable. Amplitudes are low in the center of the area for 500 Hz (Fig. 15, upper left) and much higher at the boundary of the test field. The reverse behavior of the amplitudes is visible at 600 Hz, 700 Hz, and 800 Hz. It is further assumed that these changes in the spectral amplitudes are caused by the delamination in the material. Unfortunately, a quantification of the actual area of delamination by destructive tests could not be done by the owner. Therefore, the verification of the spectral analysis still requires further investigations. A more fundamental test series was therefore performed on smaller test specimens to study the behavior of GFRC under LARS excitations.



**Fig. 15** Distribution of spectral amplitudes in the test field for different frequencies

## Investigation of Smaller Impact Specimens Made of Glass Fiber-Reinforced Polymers

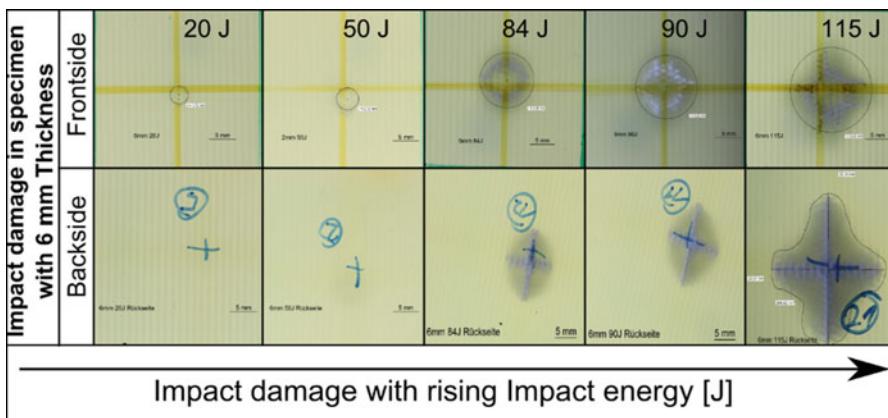
In a study conducted by Hornfeck et al. (2015), the characteristics of impact damage in specimen constructed of glass fiber-reinforced polymers were investigated to test the reliability of NDT methods. Visible turbidity, indentation depth, and bulge height were measured, and some specimens were examined with X-ray techniques (as a gold standard) to obtain detailed information on damage characteristics. Figure 16 depicts the impacted specimens sorted according to the impact energy (20 Joule to 115 Joule), where the upper row shows images of the front of the specimens and the lower row the back.

As an example of the LARS results, Fig. 17 reproduces the plotted LARS measurements of the full width at half maximum. This graph is organized according to the impact energy in the opposite way as Fig. 13, i.e., the impacts with larger impact energies are shown on the left side of the graph.

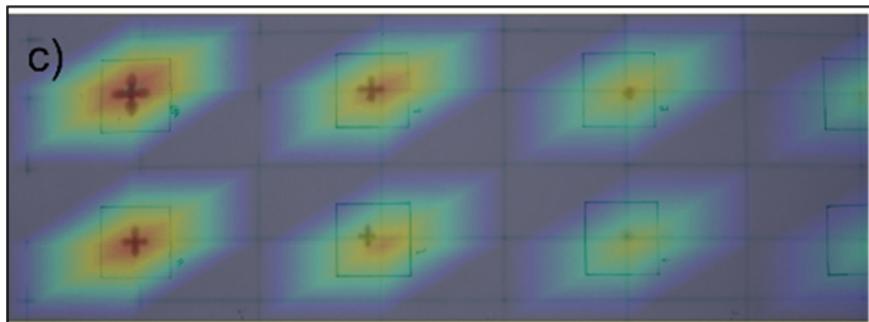
The specimens were tested with different NDT techniques including ultrasound, optical lock-in thermography, and LARS. While the small impact damage in this study was difficult to detect, if the hammer did not hit exactly its location, the authors refer to the good usefulness of the technique when applied as a first and quick inspection technique. Compared to the other techniques LARS showed, according to the authors, a relatively low technological readiness level. However, LARS has a great potential for the described applications and for many other areas.

## Applications to Test Carbon Fiber-Reinforced Polymers

In further studies, the applicability of LARS for carbon fiber-reinforced polymers (CFRP) was investigated. For this, the loading mechanism described in

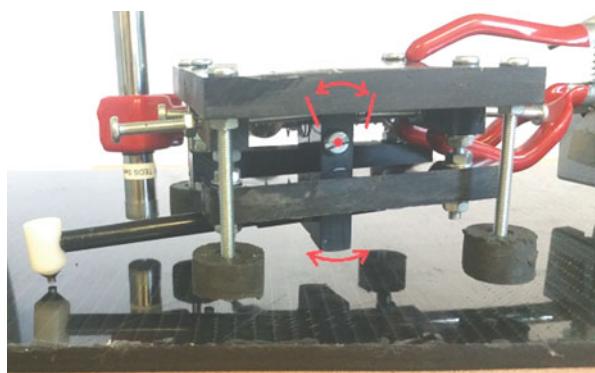


**Fig. 16** Visible impact damage at both sides according to Hornfeck et al. (2015)



**Fig. 17** Data evaluation of LARS measurements at the specimen from Fig. 16 (Hornfeck et al. 2015) using a color-coded 2D image as described in section “[Applications for Wind Turbine Blades](#)”

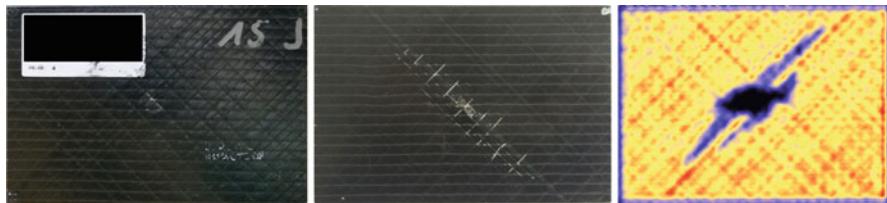
**Fig. 18** Device holding spring loaded pivoting hammer and microphone “floating” above a test piece (pivot point is marked in red) (Vandewalle 2017)



section “[Recording Equipment and Automation](#)” was refined using a spring-loaded pivoting device (Fig. 18). The system is designed in a way that multiple hammer impacts are impossible and the microphone is as close to the impact position as possible. Instead of a contact-free, “floating” system held by a tripod system, an industrial robot can be used. For all CFRP investigations described here, the hammer of Fig. 7 with a 2 mm steel tip was used to create short pulses.

Depending on the industry sector, CFRP parts are often smaller and thinner than parts made of GFRP. Accordingly, a number of small-scale test specimens with different impact-induced delaminations, flat bottom holes (FBH), and a stepped wedge were examined (Figs. 19 and 20). In the discussion that follows, all of these features are referred to as defects.

The impact energies used to create delaminations was varied from 0 to 15 Joule for CFRP plates with a thickness of 2 mm. Figure 19 shows the most damaged plate, featuring an indentation from the impactor on the front and a large cracked area on the backside. In order to define a LARS measuring grid, size and location of the



**Fig. 19** Left, front of 15 J-impacted CFRP test specimen (size:  $15 \times 10 \times 2$  mm); center, back of same specimen; right, C-scan of same piece using ultrasonic immersion technique with a Panametrics V310-5 MHz transducer

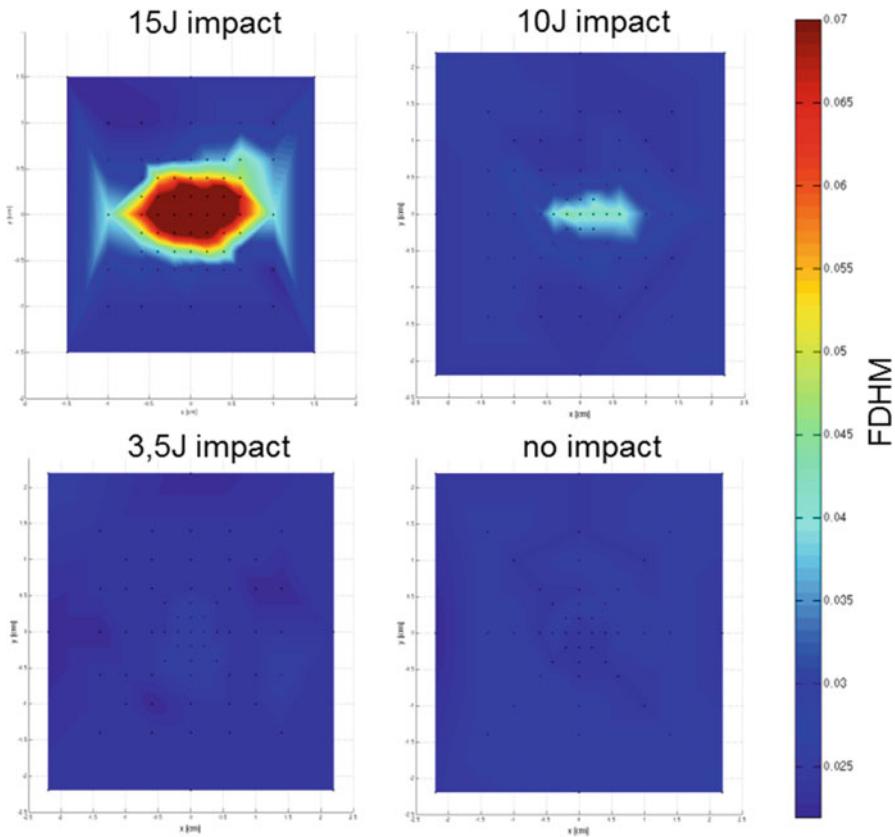
delamination were obtained by an ultrasonic immersion C-scan. Since the diagonally shaped parts of the defective area could not be detected by LARS in pretesting, analyses are focused on the indented center of the test pieces. Figure 20 shows the full duration at half maximum (FDHM) of the hammer signal for the respective area around the indentation for all four impact energies.

For both the 15 J and 10 J samples, a change in the FDHM around the center is visible, whereas the 3.5 J and the 0 J specimens show no significant features. The FDHM corresponds to the contact duration between the hammer tip and sample surface allowing two possible conclusions to be drawn. First, the contact duration is increased by the locally reduced stiffness due to the delamination, and/or second, the contact duration is increased by the indented surface where the projectile strike the sample. This seems especially possible since the diameter of the indented area is approximately five times larger than the hammer tip, causing different contact conditions. In addition, the frequency spectrum of the hammer impact in the center of the sample is considered. Figure 21 presents a comparison of the spectra of the central measuring point of all four test pieces.

Before the Fourier transformation, the ten sound signals recorded for each measuring point are correlated, normalized to the maximum of the respective hammer pulse, and averaged. It can be noted that, as expected, the spectrum contains higher frequencies if the test piece is not or only barely damaged. When the hammer hits the indentation in the center of the delaminated area, only lower frequencies occur. This difference in the frequency content is also audible with the human ear when manually testing the plates.

In order to obtain more general knowledge about the influence factors of the LARS method for CFRP, the abovementioned test piece containing flat bottom holes and a stepped wedge was used (Fig. 22).

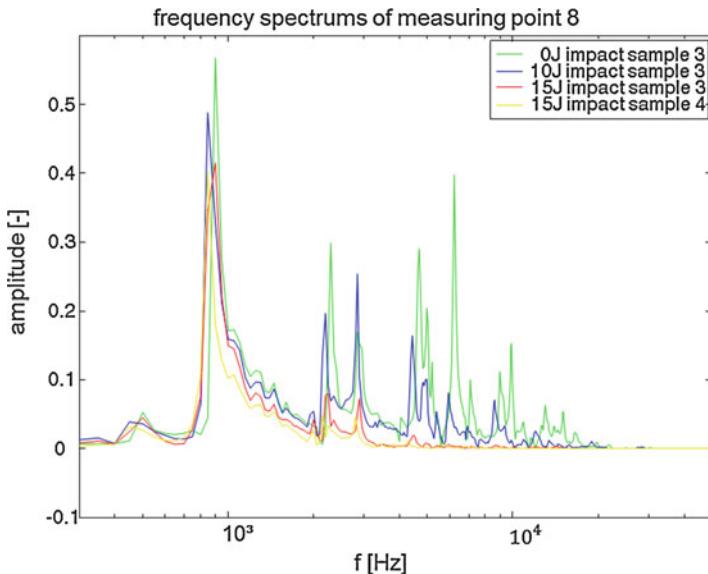
The variations of the test setup that were investigated in the study were the microphone position relative to the hammer tip and the sample surface, the type and position of the measuring grid, as well as different defect sizes and depth, represented by the FBH and stepped wedge. The study showed that the microphone should be positioned as close to the hammer tip as possible and oriented orthogonally to the sample surface. Considering the measuring grid, Fig. 23 shows two different ways of grid positioning. In the first case, measuring points located exactly



**Fig. 20** Full duration at half maximum [ms] at the center area of the impacted test pieces (Wolffhugel 2016)

over the defects are chosen to investigate the detectability of different defect depths. In the second case, a grid with regular spacing was used to simulate real testing conditions, in which the location of defects is unknown. For better visibility, the measuring points (red) are shown on the same side as the defects, whereas the actual tapping takes place on the opposite side.

In addition to the analysis of the full duration at half maximum, the sound signal energy in the time domain was chosen for analysis. Sound as well as hammer signals was normalized to the maximum of the hammer signal for better comparison. As Fig. 23 shows, the FDHM can be a good indicator to detect shallow regions in the test piece. The FBH remain undetected, presumably because they are too small and therefore do not sufficiently alter the local material properties. Also, the energy of the sound signals in the time domain does not deliver any useful information at the current state. The analysis of the distribution of spectral amplitudes in the frequency domain is subject to further investigations.



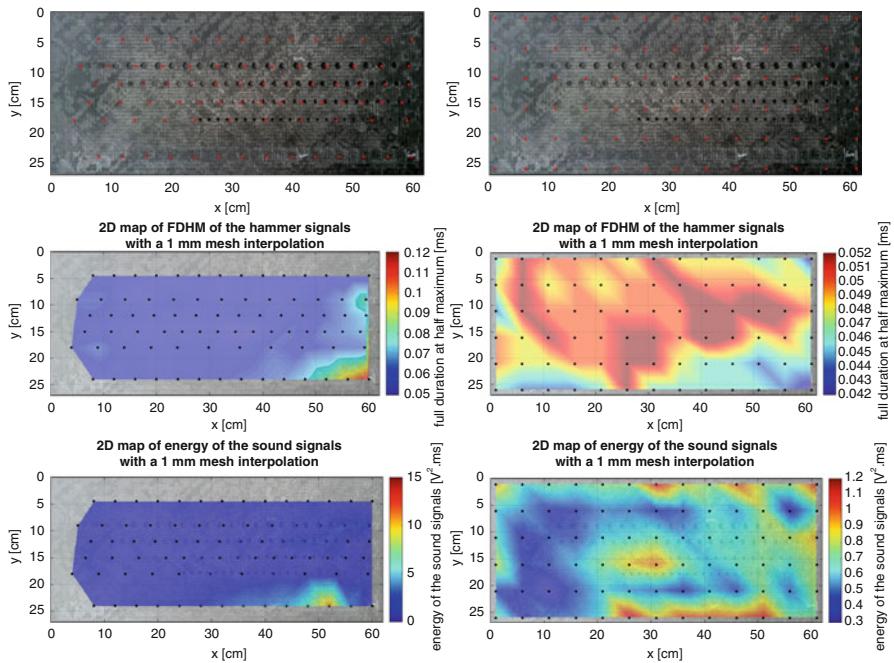
**Fig. 21** Comparison of the frequency spectra of the measuring point at the center of the impacted CFRP samples (Wolffhugel 2016)



**Fig. 22** Left, front side of FBH test specimen (size,  $620 \times 270 \times 7$  mm); right, backside of same specimen with rows of flat bottom holes of different depths, an “undamaged” reference area at the top and a stepped wedge at the bottom (Vandewalle 2017)

## Summary and Outlook

The LARS nondestructive testing method was proven to be useful as a simple and fast testing technique based on the procedures of a simple coin-tapping test (Mitsuhashi et al. 1989; Cawley and Adams 1988). The coin-tapping test is at the time being the everyday technique of experts inspecting rotor blades of wind turbines, but it is certainly not an objective test method. In particular, LARS is beneficial to identify areas with deterioration being more thoroughly investigated by other NDT techniques. This also leads to a way of escalation of NDT techniques and a quality control of components being more robust and efficient. Automation using, e.g., a robot system render possibilities for fast and automatic applications of LARS



**Fig. 23** Top left, measuring grid over the defects; top right, measuring grid with regular spacing, tapping on opposite side; center left, map of FDHM distribution on defect grid; center right, map of FDHM distribution on regular grid; bottom left, map of time domain signal energy distribution on defect grid. Bottom right: map of time domain signal energy distribution on regular grid (Vandewalle 2017; Jatzlau et al. 2016)

even in large structures. Defect types that can be detected are delaminations, cracks parallel to the surface, and voids.

Most of the applications reported so far are in the field of fiber-reinforced polymers (GFRC, CFRC) but other applications are certainly possible. One of them is based on the analogy between LARS and the impact-echo technique (Carino et al. 1986; Sansalone and Streett 1997). Conventional impact-echo (IE) techniques are used to investigate concrete structures using a steel ball excitation and a broadband sensor coupled to the surface at a certain point (Schickert et al. 2009). IE data processing is normally done in the frequency domain to observe the spectrum for amplitude maxima related to the depth of reflectors inside the concrete specimen. A more comprehensive theoretical evaluation of this technique by Gibson and Popovics (2005) showed evidence that the “resonance peak” is formed by the frequency of the S1 lamb mode with zero group velocity. Groschup and Grosse (2015) reported later about the detection of delaminations (bond failures) in concrete and in particular highway pavements and decks using a fast impact-echo scanner. They used an array of MEMS microphones instead of conventional condenser microphones that generally allow for a contact-free measurement of the structural response on impacts. This step brought impact-echo and LARS closer together.

MEMS microphone arrays could also be an option to improve LARS further. In the same paper by Groschup and Grosse (2015), a way to use numerical simulations improving the setup and the emitter and recorded frequencies is presented that could be another way to improve LARS. Numerical simulations could help identify the best emitter/receiver combination and the measuring grid (distance between scanning lines and individual tapping points) and so to optimize LARS for a particular application.

For advances on the signal processing side, audio signal processing techniques using psychoacoustic algorithms are promising (Andreisek et al. 2015, 2016). They can emulate the abilities of the human audio perception as it is used in the conventional coin-tapping test. This would have the advantage of making quality control independent from the experience of a user and his hearing capabilities.

---

## References

- Abrate S (2005) Impact on composite structure. Cambridge University Press, Cambridge
- Andreisek G, Korthals D, Grosse CU, Seeber BU (2016) The virtual tap test – a training system for wind turbine rotor blade inspectors. In: Proceedings of the 19th world conference on non-destructive testing. WCNDT 2016, Th.4.E.3. pp 1–7
- Andreisek G, Grosse CU, Seeber BU (2015) Attribute zur Beschreibung akustischer Unterschiede von Fehlstellen an Rotorblättern von Windenergieanlagen. Fortschritte der Akustik – DAGA '15. Dt. Ges. f. Akustik, Berlin, pp 513–515
- Carino NJ, Sansalone M, Hsu NN (1986) A point source–point receiver technique for flaw detection in concrete. J Am Conc Inst 83(2):199–208
- Cawley P, Adams RD (1988) The mechanics of the coin-tap method of non-destructive testing. J Sound Vib 122:299–316
- Gibson A, Popovics JS (2005) Lamb wave basis for impact-echo method analysis. ASCE J Eng Mech 131:438–443
- Groschup R, Grosse CU (2015) MEMS microphone Array sensor for air-coupled impact-Echo. Sensors 15:14932–14945. <https://doi.org/10.3390/s150714932>
- Grosse CU, Goldammer M, Grager JC, Heichler G, Jahnke P, Jatzlau P, Kiefel D, Mosch M, Oster R, Sause MGR, Stöbel R, Ulrich M (2016) Comparison of NDT Techniques to Evaluate CFRP – Results Obtained in a MAIzfp Round Robin Test. In: Proceedings of the world conference on NDT. German Society of NDT, München
- Ham S, Popovics J (2015) Application of micro-electro-mechanical sensors contactless NDT of concrete structures. Sensors 15:9078–9096
- Hertz H (1881) Über die Berührung fester elastischer Körper. Journal für reine und angewandte Mathematik 92:156–171
- Hornfeck C, Geiss C, Rücker M, Grosse CU (2015) Comparative study of state of the art nondestructive testing methods with the local acoustic resonance spectroscopy to detect damages in GFRP. J Nondestruct Eval 34(2):1–14
- Jatzlau P, Müller M, Grosse CU (2016) Identification of flawed CFRP samples using local acoustic resonance spectroscopy (LARS). In: Proceedings of the 19th world conference on non-destructive testing, WCNDT 2016, NDT.net. 8p
- Jüngert A (2010) Untersuchung von GFK Bauteilen mit akustischen Verfahren am Beispiel der Rotorblätter von Windenergieanlagen. Ph.D. Thesis, University of Stuttgart, 179p
- Jüngert A, Grosse CU, Krüger M (2013) Local acoustic resonance spectroscopy (LARS) for glass Fiber-reinforced polymer applications. J Nondestruct Eval 33(1):23–33

- Landau LD, Lifschitz EM (1989) Lehrbuch der Theoretischen Physik, Bd. VII Elastizitätstheorie, 6. Aufl. Akademie-Verlag (Berlin), 223p
- McLaskey G, Glaser SD (2010) Hertzian impact: experimental study of the force pulse and resulting stress waves. *J Acoust Soc Am* 128(3):1087–1096
- Mitsuhashi K, Jyomuta C, Oka F, Nishikawa H (1989) Method and apparatus for impact-type inspection of structures. US Patent No 05,048,320
- Müller M (2015) Möglichkeiten und Grenzen von Resonanzanalyseverfahren zur zerstörungsfreien Prüfung von Faser-Kunststoff-Verbunden. Bachelor Thesis, Technical Univ. of Munich, Chair of Non-destructive Testing, 145p
- Narr A (2017) Einfluss von impact-Schäden bei der experimentellen Modalanalyse von CFK-Platten. Bachelor thesis, Technical University of Munich, chair of non-destructive testing, 131p
- Olympus (2010) Bondmaster 1000+ – multimode adhesive bond testing application guide. Olympus, Tokyo
- Sansalone MJ, Carino NJ (1986) Impact-echo: a method for flaw detection in concrete using transient stress waves. In: National Bureau of Standards. NISEE, Berkeley
- Sansalone MJ, Streett W (1997) Impact-Echo: nondestructive evaluation of concrete and masonry. Bullbrier Press, Jersey Shore
- Schickert M, Neisecke J, Brameshuber W, Colla C, Flohrer C, Gardei A, Grosse CU, Krause M, Krogel O, Krüger M, Willmes M (2009) DGZfP Merkblatt B11 – a guideline describing fundamentals and applications of the impact-Echo method, international symposium Non-Destructive Testing in Civil Engineering (NDT-CE), ISBN: 978-2-7208-2542-5, Nantes, pp 807–811
- Shi X, Polycarpou AA (2005) Measurement and modeling of normal contact stiffness and contact damping at the Meso scale. *J Vib Acoust* 127:52–60
- Vandewalle N (2017) Local acoustic resonance spectroscopy. Non-destructive defect detection for carbon fiber reinforced composites, research internship report (PRE) at TU Munich, ENSTA ParisTech, 43p
- Wolffhugel T (2016) Local acoustic resonance spectroscopy. Research internship report (PRE) at TU Munich, ENSTA ParisTech, 67p
- Zhu JY, Popovics JS (2007) Imaging concrete structures using air-coupled impact-echo. *ASCE J Eng Mech* 133:628–640



# Nonlinear Resonant Acoustic Spectroscopy 10

Bart Van Damme and Koen Van Den Abeele

## Contents

Nonlinear Elasticity and Nondestructive Testing .....	296
Introduction .....	296
Classical and Nonclassical Nonlinear Elasticity .....	297
Nonlinear Behavior of Damaged Materials .....	301
Nonlinear Reverberation Spectroscopy .....	304
Applications of NRS .....	305
Application to Thermally Loaded CFRP .....	305
Detection of Fatigue Damage in CFRP and Steel .....	314
Discussion .....	318
Conclusion .....	320
References .....	320

## Abstract

Ultrasound nondestructive evaluation methods are popular since they are noninvasive, can be done by a trained technician, and can be used *in situ*. Pulse echo measurements with a single transducer and more advanced material examination using phased arrays allow for the detection of cracks and cavities in homogeneous materials. However, some types of damage remain invisible to traditional ultrasound nondestructive testing (NDT). Early stage fatigue damage, closed cracks, or delaminations in composites are typically difficult to discern. This chapter deals with the assessment of nonlinear wave distortion due to the presence of damage. In nonlinear elastic materials, resonance frequencies depend on the

---

B. Van Damme (✉)

Empa, Swiss Federal Laboratories for Materials Science and Technology, Dubendorf, Switzerland  
e-mail: [bart.vandamme@empa.ch](mailto:bart.vandamme@empa.ch)

K. Van Den Abeele

Physics, Kulak Kortrijk Campus, Kortrijk, Belgium  
e-mail: [koen.vandenabeele@kuleuven.be](mailto:koen.vandenabeele@kuleuven.be)

excitation amplitude. Nonlinear reverberation spectroscopy (NRS) exploits this small frequency shift in the ringing of a sample that was harmonically excited at resonance. Two successful applications are described. First, NRS was used to quantify thermal damage of carbon fiber reinforced polymers (CFRP) samples. The nonlinear parameters are much more sensitive to the microdamage than linear properties, such as a change of the Young's modulus or the damping coefficient. Moreover, the NRS results correlate well with optically gathered crack density values, and they can be modeled using a hysteretic elastic constitutive equation. The second example is the detection of early fatigue damage. A single closed crack in a steel sample can be exposed and even located when combining the NRS results with a finite element modal analysis.

---

## Nonlinear Elasticity and Nondestructive Testing

### Introduction

Traditional techniques for nondestructive testing (NDT), acoustical as well as others, are based on a contrast between the damaged zone and the intact material. This contrast can be a change in mechanical impedance, causing reflections and absorption of elastic waves (echography, sonar, acoustic scans) or a change in electromagnetic impedance (eddy current spectroscopy) ([Mix 2005](#)). When using ultrasound methods to assess materials, scale limitations arise. In general, the defect detection sensitivity increases for smaller wavelengths ([Schmerr 2016](#)). However, high frequency pulses are highly attenuated thereby limiting the application field. Sub-wavelength sized cracks can be detected by analyzing ultrasonic array signals, which requires costly equipment ([Zhang 2016](#)). Even then, there are certain cases where these techniques are not always successful. A single closed crack for instance establishes a particular problem for the strength of a material, but is nearly invisible for traditional acoustic techniques since an incoming wave is not reflected. A possible approach to overcome this problem is using nonlinear wave propagation properties of damaged materials ([Guyer and Johnson 2009](#)).

The phenomenon of nonlinear elastic behavior has been discovered as early as the eighteenth century, but its practical repercussions were only addressed halfway the twentieth century ([Hamilton 1986](#)). Whereas most homogeneous solid materials, like metals, show a linear stress-strain relationship, a lot of anomalies are known ([Guyer and Johnson 2009](#); [Nazarov and Radostin 2015](#)). For large stresses in the elastic regime, the strain of most materials deviates from the linear Hooke's law. Other materials like sandstone show a strong nonlinear stress-strain relationship at much lower elastic stresses and even hysteresis when a stress cycle is applied. These effects have been investigated thoroughly over the past decades, and models were developed to describe the particular nonlinearity of different classes of materials.

Linear materials with local damage or with a distribution of microcracks show a higher nonlinear elastic behavior than an undamaged reference material ([Van Den Abeele et al. 1997](#); [Nagy 1998](#); [Kim et al. 2011](#)). In some cases where classical NDT

techniques fail, the measurement of a slight variation in the nonlinear material parameters can reveal early damage. The idea behind this concept is the fact that small cracks or delaminations are *activated* by the applied dynamic stress field, resulting in a response signal that is different from the expected response in the linear regime. The larger the activation energy, the larger the nonlinear response. The observation of this amplitude dependency will be used as an alternative for the classically used and widely spread commercial techniques available. Techniques based on nonlinear material behavior are commonly called nonlinear elastic wave spectroscopy (NEWS).

In this chapter, a short introduction to some widely used nonlinear material models is given first. The section is restricted to models necessary for the comprehension of the nonlinear NDT method. After an outline of the models, some examples of measurable footprints, produced by acoustic waves passing through nonlinear elastic zones, are given. The NEWS technique described in this chapter, nonlinear reverberation spectroscopy, are based on the detection of these nonlinear footprints.

## Classical and Nonclassical Nonlinear Elasticity

### Classical Nonlinearity

The relation between stress and strain determines the elastic modulus  $K$ , which is defined as

$$K = \frac{d\sigma}{d\epsilon}, \quad (1)$$

in which  $\sigma$  is the applied stress and  $\epsilon$  is the strain. In linear materials, the elastic modulus is a constant  $K_0$ , resulting in Hooke's law

$$\sigma = K_0\epsilon. \quad (2)$$

The simplest case of nonlinear behavior can be described using a series expansion for the stress as a function of strain (Landau and Lifshitz 1959). This behavior can be described analytically as

$$\sigma = K_0\epsilon + K_0\beta^*\epsilon^2 + K_0\delta^*\epsilon^3 + \dots \quad (3)$$

which can be rewritten using Eq. 1 as an elastic modulus

$$K = K_0(1 + \beta\epsilon + \delta\epsilon^2 + \dots). \quad (4)$$

The perturbations  $\beta\epsilon$  and  $\delta\epsilon^2$  are typically much smaller than 1, and they describe the nonlinear hardening or softening of the material when a stress is applied. These perturbations are a function of the second and third (in the case of  $\beta$ ) and even fourth (in the case of  $\delta$ ) order elastic constants of the material. When a sine wave is passing

through a medium and meets a zone governed by this constitutive relation, the original sine will be deformed and harmonics will be created. In case of first-order nonlinearity only, meaning that  $\delta = 0$ , even and odd harmonics show up in the resulting signal. When only second-order nonlinearity is present ( $\beta = 0$ ), only odd harmonics will be created.

A material whose deformation is described by this relationship is called a *classical nonlinear* material. In linear elastic materials such as metals or Plexiglas without any damage, this nonlinear behavior is normally very small and difficult to measure. The sources of this nonlinearity however are well described and can be linked to deformations and dislocations at the atomic scale (Landau and Lifshitz 1959; Hamilton 1986). Classical nonlinearity is therefore also referred to as *atomic nonlinearity*. Since any material contains a certain amount of atomic defects, the nonlinear parameters will never be zero. It is important to know the natural non-linearity level in order to quantify the influence of developing damage.

## **Nonclassical Nonlinearity: Hysteresis and the Preisach-Mayergoyz Approach**

Micro-inhomogeneous materials like sandstone or concrete exhibit an even more complex demeanor when stress is applied. It has been known for a long time that rocks show a hysteretic behavior, comparable with the well-known hysteresis in ferromagnetic materials. When a cyclic stress is applied, the resulting strain exhibits a loop because of a different strain response when the material is compressed or stretched (Guyer and Johnson 1999). An example for sandstone is shown in Fig. 1.

As a result, the elastic modulus has discontinuities when the applied stress changes from positive to negative. In general, this can be described by

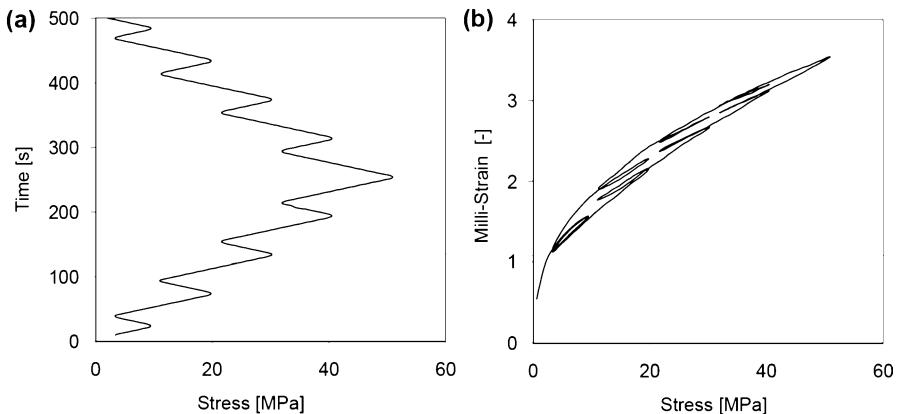
$$K = K(\epsilon, \dot{\epsilon}), \quad (5)$$

the time derivative  $\dot{\epsilon}$  describing the dependency of the modulus on the direction in which the strain is varying, and its rate of change. When the strain diminishes, the value for  $K$  is different than when the strain increases. This dependency of the change of strain in time implies a memory effect of the hysteretic behavior. The memory shows up in the hysteretic curve in Fig. 1, where all inner loops arrive at the exact same place as where they started. This phenomenon is called endpoint memory, and it has been observed in a variety of materials.

More explicitly, the hysteretic stress-strain relation for a cyclic stress function can be described to first order by the equation

$$K(\epsilon) = K_0(1 - \alpha[\Delta\epsilon + \epsilon \operatorname{sign}(\dot{\epsilon})]), \quad (6)$$

where  $\Delta\epsilon$  is the amplitude during the last period and  $\alpha$  is a constant describing the hysteretic behavior. The consequences of this elastic law for harmonic waves were described in a quasi-analytical way by Van Den Abeele et al. (1997). The most important result is that in this case of nonclassical nonlinear elasticity only odd harmonics are generated and that their amplitudes are proportional to the square of the fundamental strain amplitude.

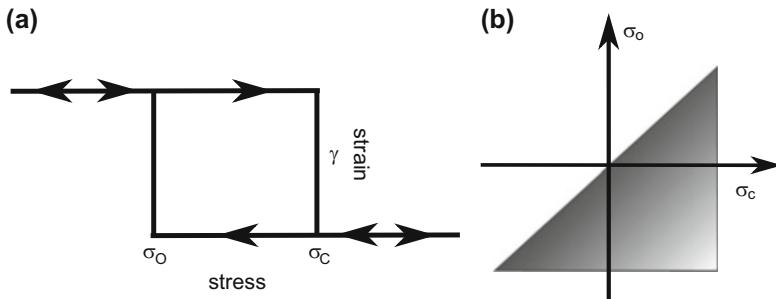


**Fig. 1** Example of hysteretic behavior of Serena sandstone. The stress-strain relation (b) is shown for an applied stress protocol (a). Internal closed loops are visible

The physical explanation of this macroscopic behavior is commonly attributed to friction between the grains of a material (Walsh 1966; Mason 1969; Mavko 1979), although Sharma and Tutuncu (1994) found some discrepancies in this theorem and they added adhesion to this model. Most of these models however are purely descriptive, and very little analytical work is done (Pecorari 2004). All models however are based on the mesoscopic of materials. Mesoscopic refers to the scale between the nanolevel (atomic level, cf. classical nonlinearity) and the macroscopic properties, in this case the behavior of grains in stone or metals with a size varying from 0.1 to 10  $\mu\text{m}$ . Since the bonds between the grains typically behave completely different than the material inside the grains, this results in nonclassical stress-strain relationships (Ten Cate and Shankland 1996; Darling et al. 2004).

A phenomenological approach to describe hysteresis is the use of the Preisach-Mayergoyz (PM) framework. This model was originally developed for the description of hysteresis in ferromagnetic materials but can also be used to deal with static and dynamic mechanical hysteresis (McCall and Guyer 1994; Guyer et al. 1995). The basis of this approach is to represent the material as an ensemble of features called hysterons. These entities can be either in an open or closed state. The stress value for which a hysteron opens ( $\sigma_o$ ) or closes ( $\sigma_c$ ) can be different for each individual hysteron. The strain caused by the opening of a single hysteron is assumed to be a constant value  $\gamma$ . The distribution of all hysterons can be represented in a space defined as a density of entities as function of the switching values  $\sigma_o$  and  $\sigma_c$ . This space is called the PM-space and the PM-space density  $\rho_{PM}(\sigma_o, \sigma_c)$ . This is represented schematically in Fig. 2.

The PM space, and the exact distribution of hysterons in particular, is usually applied as a purely mathematical tool to describe and model the hysteretic properties of a material. Some researchers however are investigating a universal theory about the physical grounds of the exact distribution of hysterons in PM space. Work is



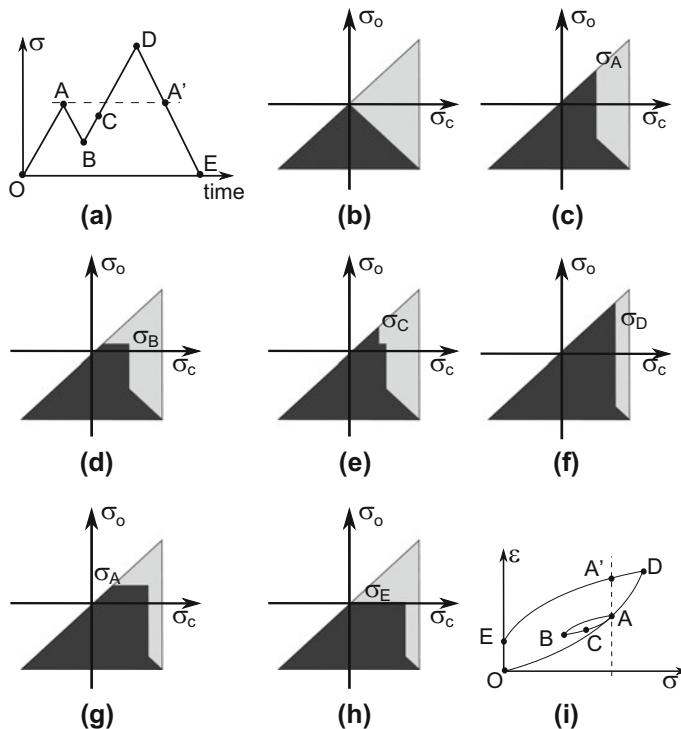
**Fig. 2** Hysteron (a) and PM space distribution (b)

done by Aleshin and Van Den Abeele (2005, 2007, 2009) in the case of Hertzian friction between microscale material cells, but in most cases, a simple analytical distribution is used to approximate the reality. Some possibilities found in literature include a uniform distribution or an exponential distribution, scaled in such a way that the entire volume is equal to one. Experimental work on the dependency of the PM distribution on external factors such as humidity was done on concrete by Van Den Abeele et al. (2002).

The approach to calculate the strain from the stress is completely different than in the case of classical (nonlinear) elasticity and is only based on the amount of hysterons that are open and closed. The total strain of the medium is calculated as the amount of open hysterons, multiplied by the hysteron strain value  $\gamma$ . The amount of open and closed hysterons not only depends on the current stress and stress rate but on the entire history of the stress, on the initial states of the hysterons, and on the distribution of the hysterons in the PM space. The following example explains the procedure to calculate the strain for a certain stress cycle shown in Fig. 3a.

The PM space is represented in a plane defined by  $\sigma_c$  on the horizontal and  $\sigma_o$  on the vertical axis. The distribution of the hysterons in the PM space is situated in a triangular zone defined by the first bisector (for which  $\sigma_c = \sigma_o$ ) and a horizontal and vertical line defining the maximum values of the opening and closing stresses. In theory, the vertical and horizontal limits of this zone can be infinite, but for practical calculations mostly finite borders are chosen. An example of the evolution of closed and open hysterons due to a predefined stress cycle is shown in (Fig. 3). In every situation, one single continuous line divides the set of open (on the upper right side) and the closed (on the lower left side) hysterons.

A purely hysteretic model does not take into account any classical elasticity. Most materials showing hysteretic behavior however can best be described using a combination of both hysteretic and classic elasticity. In for example sandstone, the sand grains can be seen as elastic building blocks of the material, whereas the bonds between the grains will behave in a hysteretic way. This model was, for instance, used in the simulation of resonant bars with localized damage (Van Den Abeele et al. 2004b).



**Fig. 3** Example of the evolution in PM space when applying the stress cycle shown in (a). The resulting strain can be seen in (i). The dark zone matches with closed hysteresons, the lighter zone stands for open hysteresons

## Nonlinear Behavior of Damaged Materials

It is well known that the nonlinear elastic behavior of materials increases in the presence of damage, such as cracks (Nagy 1998; Van Den Abeele et al. 2000a, b). Early experiments using nonlinear methods for detecting small cracks in various materials such as steel, concrete, or carbon were performed by Antonets et al. (1986), Shkolnik (1993), Korotkov and Sutin (1994), Sutin et al. (1995). The exact physical basis of nonlinearity generation by cracks however is not completely understood, and a variety of models have been developed.

When applying a shear stress on a crack, both sides will slide. Mesoscopic friction models describe a stick-slip mechanism that is highly nonlinear. Theoretical investigations were done by Lawn and Brian (1998), Pecorari (2003), and more recently by Aleshin and Van Den Abeele (2007). Pecorari proved the presence of classical and hysteretic nonlinear effects, resulting in even and odd harmonics, respectively, whereas Aleshin related the contact of rough surfaces to a description of PM space distributions.

The nonlinearity of a sample is directly related to the damage state (Sutin et al. 2003; Gliozzi et al. 2006; Zumpano and Meo 2007); thus, the knowledge of the nonlinearity parameters can be used to detect and quantify the damage, and even to distinguish the type of damage from its specific nonlinear behavior (classical or hysteretic). Waves passing through the nonlinear elastic zones are distorted in a way that can be distinguished macroscopically. Most of these effects are based on amplitude dependent frequency changes and the generation of new frequencies in a sample when an external monochromatic sound wave is applied.

**Harmonics generation** The most direct effect of a nonlinear stress-strain relationship to assess damage is the generation of harmonics. Since the nonlinear contributions to the elastic modulus deform any sinusoidal wave traveling through a medium, harmonics will be generated. This approach has been successfully applied for cracks in aluminum samples (Buck et al. 1978; Morris et al. 1979) and carbon fiber-reinforced polymers (Kawashima et al. 2006; Solodov and Busse 2007).

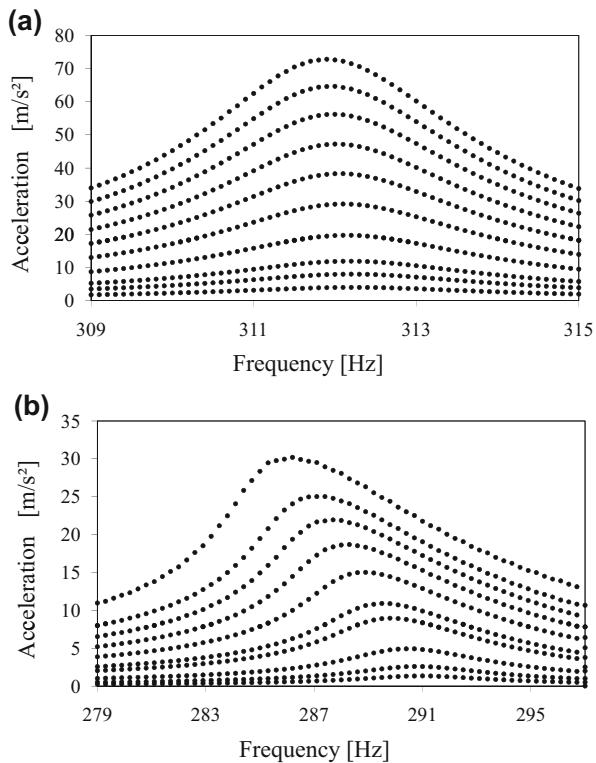
Although the generation of harmonics is of large theoretical interest, experiments are often disturbed by nonlinear contributions from signal amplifiers that also deform sinusoidal signals. When using high amplitude excitation, it is therefore not easy to distinguish the harmonics produced by the damaged sample and the electronics used. Reference measurements on linear samples using the same electronic setup can be performed in order to filter out the erroneous effects.

**Nonlinear wave modulation** Another well-known nonlinear phenomenon is the mixing of two independent waves with frequencies  $f_0$  and  $f_1$  ( $f_0 > f_1$ ) when passing through a nonlinear medium. The waves interact, causing amplitude modulation effects. The resulting signal will not only consist of the original frequencies and their harmonics, but sideband frequencies  $f_0 \pm f_1$ ,  $f_0 \pm 2f_1$ , ... will also be created (Van Den Abeele et al. 2000b). The presence and amplitude of the sidebands depend on the type of nonlinearity.

The first experiments based on this particular nonlinear feature (nonlinear waveform modulation spectroscopy or NWMS) were performed in the 1990s by Sutin and Donskoy (1998). The technique is mostly used to detect cracks, for instance, by Courtney et al. (2008) and Zagrai et al. (2008), and for the evaluation of multilayered structures (Antonets et al. 1986). Kazakov et al. (2002) developed a way of localizing damage, comparable to the method described using harmonics detection.

Extensive work was done by Van Den Abeele et al. (2000b) in developing experiments to detect single cracks in Plexiglas and automotive parts. The authors were able to distinguish higher order modulation effects using a high frequency transducer and a low frequency shaker, and they compared the results with those of the harmonics detection. In the same paper, a method is described that uses a short impact (transient excitation) on the sample to produce a broadband low frequency signal consisting of the eigenfrequencies of the material. This technique is called impact modulation.

**Fig. 4** Resonance curves of an intact (a) and damaged (b) slate beam sample measured for increasing driving amplitudes. (Reproduced from Van Den Abeele et al. 2000a, with the permission of the American Society for Nondestructive Testing, Inc.)



The main advantage of modulation effects over harmonics detection is the possibility of using independent wave sources for both signals. If there is no electronic connection between both sources, modulation frequencies can only be generated by the nonlinear material properties since no other nonlinear sources are present.

**Amplitude-dependent eigenfrequency** Since the eigenfrequencies of a material directly depend on the stiffness, an amplitude-dependent elastic modulus will result in an amplitude dependency of the eigenfrequencies. This phenomenon has, for instance, extensively been studied in inherently nonlinear materials, such as rocks (Johnson et al. 1996, 2004; Johnson and Sutin 2005) and for many other materials (Van Den Abeele et al. 2000a, 2006; Van Den Abeele and De Visscher 2000). The resonance frequency in sandstone shifts to lower values for increasing amplitudes. The nonlinear effect increases in damaged materials. An example is given in Fig. 4 for an undamaged and a damaged slate beam (Van Den Abeele et al. 2000a).

The authors developed a method of detecting damage based on the relative frequency shift, which they called Single Mode Nonlinear Resonance Acoustic/Ultrasound Spectroscopy (SiMoNRAS/SiMoNRUS). Resonance peaks are measured doing a step-by-step sweep over a fixed frequency range for increasing amplitudes.

This frequency domain method has shown to be effective for detecting damage in a variety of materials (Van Den Abeele et al. 2006), but the method is time consuming.

**Slow dynamics** It has been observed that the elastic properties of a material temporarily change after applying an elastic load. It is clear from the previous section that the resonance frequency of a sample decreases for larger amplitudes. In addition, after performing a SiMoNRAS experiment on sandstone at relatively high amplitude, a new sweep at low amplitude will result in a lower resonance frequency than was found in the original low amplitude sweep, which means the material has been softened by the previous high amplitude sweep. If the same frequency sweep at low strain amplitude is repeated, the resonance frequency shifts back toward its original value, indicating a relaxation in time of the material. This phenomenon is called slow dynamics.

Since slow dynamics is a proof of mesoscale nonlinearity, it can again be used for the detection of hysteretic features. Bentahar et al. (2006) and Johnson and Sutin (2005) compared fast (SiMoNRAS) and slow dynamic techniques for damage detection in concrete, both in a numerical way and experimentally. The authors found that damaged materials need much more time for the relaxation after applying a load, and the relative frequency change is much higher than in intact samples.

---

## Nonlinear Reverberation Spectroscopy

Nonlinear resonance techniques investigate the conduct of objects when they are excited by increasing amplitudes. Generally, a single resonance mode of the object with its associated frequency is selected. In SiMoNRUS (Johnson et al. 1996; Van Den Abeele et al. 2000a), the object is subjected to a frequency sweep around this resonance frequency at constant excitation amplitude. The true resonance characteristics, frequency and damping (or quality factor), are then analyzed from fits of the resulting frequency response amplitude (resonance curve). Multiple resonances of complex objects can be exploited for a better estimation of the linear elastic properties.

The experiment can be repeated for increasing excitation amplitudes to analyze the nonlinearity of the object. Intact materials show no change in the resonance characteristics, whereas damaged materials generally show a decrease in the resonance frequency with amplitude (nonlinear softening) and an increase in the damping factor ( $\pi f_0/Q$ , with  $Q$  as the quality factor) due to nonlinear attenuation (Van Den Abeele and De Visscher 2000; Van den Abeele et al. 2002; Johnson and Sutin 2005).

Nonlinear reverberation spectroscopy (NRS) is the time domain analogy of SiMoNRUS. In NRS, a sample is excited using a constant excitation at a single frequency for a certain period of time. The frequency is chosen in the neighborhood of one of the resonance frequencies of the sample. After a number of cycles, sufficient to reach a steady-state response, the continuous wave excitation is stopped, say, at  $t = t_0$ . The reverberation response of the sample is measured from  $t_0$  to  $t_1$  and

stored for analysis. The recorded signal is a typical decaying time signal, with large amplitudes near  $t_0$  and smaller amplitudes near  $t_1$ . This response is analyzed using a successive fitting of an exponentially decaying sine function,

$$A_k e^{-\xi_k t} \sin(2\pi f_k t + \phi_k) \quad (7)$$

to small time windows (approximately 20 cycles). Here,  $A_k$  denotes the amplitude,  $\xi_k$  is the decay parameter,  $f_k$  is the frequency, and  $\phi_k$  is the phase of the signal in the  $k$ th window. This allows the creation of a parametric plot of the true resonance frequency  $f_k$  and of the decay parameter  $\xi_k$  as a function of the amplitude  $A_k$ , thereby providing information on the occurrence of nonlinearity. If the material is linear, the frequency in different windows of the reverberation signal remains constant. If the material is nonlinear, the frequency in the reverberation signal gradually increases with decreasing amplitude and thus with time, in agreement with the nonlinear softening effect on the modulus due to the presence of nonlinearity (Guyer and Johnson 1999; Guyer et al. 1998).

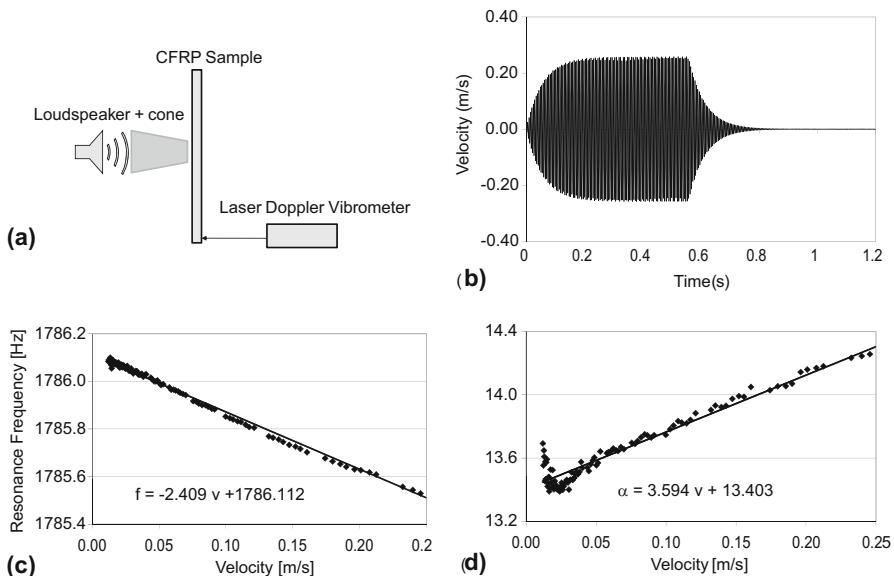
SiMoNRUS and NRS can be performed in a fully noncontact mode by means of a loudspeaker as exciter and a laser Doppler vibrometer for the response measurement. The schematic setup and a typical NRS response and analysis of the data for one of the samples considered in section “[Application to Thermally Loaded CFRP](#)” (thermally shocked CFRP) can be found in Fig. 5a, b. The amplitude dependence of the resonance frequency and damping are clear markers of the nonlinear material behavior. When the amplitudes are recalculated in terms of strain (see section “[NRS Results](#)”), a NRS nonlinearity parameter can be deduced from the proportionality relation as the slope of the relative change.

## Applications of NRS

### Application to Thermally Loaded CFRP

#### CFRP and Heat Damage

Carbon fiber-reinforced polymers (CFRPs) are commonly used in the aircraft construction industry. It is expected that the next generation of airplanes will consist of more than 60% of composite structures (Taylor 2011). Even though composite materials hold important advantages over aluminum, CFRP is also prone to various degradation mechanisms. The exposure to heat, for instance, induces chemical and microstructural changes affecting the mechanical behavior of the composite laminate, even at moderate temperatures (Matzkanin 1999). Traditional nondestructive quality control techniques are often limited in their capabilities of detecting and characterizing subtle changes in the material properties associated with heat damage. A review of the mechanisms of heat damage in composites and a state of the art of nondestructive evaluation (NDE) techniques currently used to evaluate heat damage can be found in a comprehensive article by Matzkanin (1999).

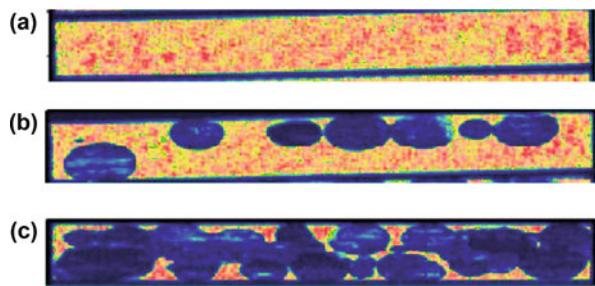


**Fig. 5** NRS experimental setup (a) and typical results: (b) Full recorded signal; (c) Analysis of instantaneous resonance frequency versus particle velocity amplitude for a CFRP sample shocked at 300 °C for 30 min; (d) Analysis of instantaneous damping characteristic versus particle velocity amplitude for the same sample. (Reproduced from Van Den Abeele et al. 2009, <https://doi.org/10.1121/1.3184583> with the permission of the Acoustical Society of America.)

Studies have shown that thermal degradation is typically matrix dominated since by the time fiber properties such as tensile strength and modulus are affected, all other mechanical integrity is lost. Mechanical metrics such as compressive, shear, and flexural strength and stiffness properties are believed to be the most sensitive properties for use in the early detection of thermal degradation, as opposed to nonmechanical parameters such as thermal and dielectric properties. Most of the work reported in the literature dealing with NDE for heat damage in composites deals with the following five methods: thermal (IR), ultrasonics, acoustic emission, dielectric properties, and radiography. These methods, while being readily available and well developed, are limited in their capabilities to detect and characterize the changes in composite material properties associated with heat damage. For instance, the detectability threshold of heat damage (1 h exposure at temperatures 200–300 °C in unidirectional AS4-8552 CFRP laminates using conventional ultrasonics (immersed transmission C-scan imaging at 5 MHz)) was found at 290 °C (Hyllengren 2001). This can be seen in the C-scan examples in Fig. 6. Nevertheless, the measured value of the interlaminar shear strength for the same type of samples changed from 121 MPa for nonexposed samples to 114 MPa when exposed at 200 °C, to 84 MPa for 285 °C, and to 43 MPa for samples exposed at 300 °C for the duration of 1 h.

Most traditional NDE techniques are capable of detecting physical anomalies such as cracks and delaminations. However, to be effective for thermal degradation, they must be capable of detecting initial heat damage, which occurs at a microscopic

**Fig. 6** Comparison of linear C-scan results for thermally loaded CFRP bars. The samples were exposed to 285 °C (a), 290 °C (b), and 300 °C (c) for 1 h. Delaminations (in dark blue) are only visible in the last two samples



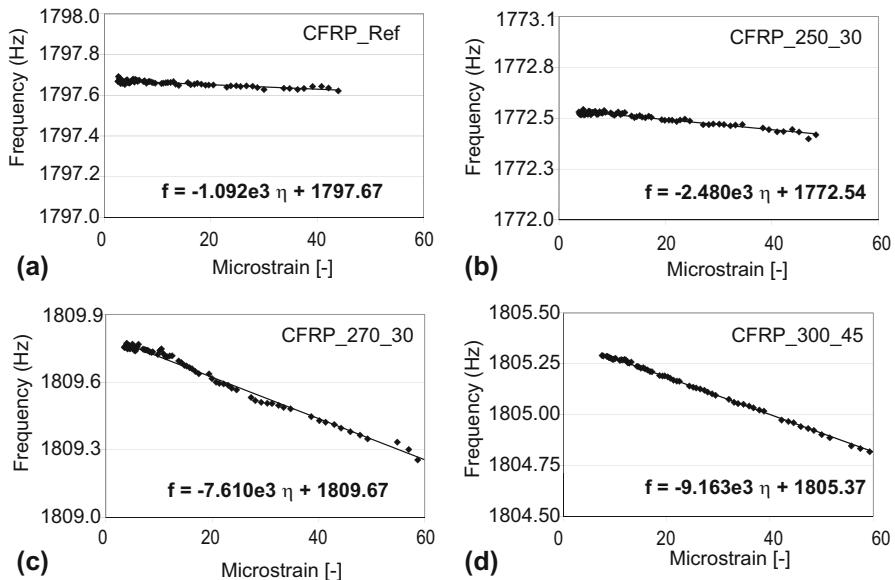
scale. Review of the literature from more recent years indicates that a vast number of NDE methods are currently under development and show various degrees of promise for characterizing heat damage in composites.

### NRS Results

We examined a set of heat damaged composite laminate samples using the above described NRS technique and quantified their NRS nonlinearity parameters as function of the heating temperature and exposure time. The set of 21 CFRP (AS4/8552 quasi-isotropic lay-up) samples consisted of one reference sample, which was left unexposed, and 20 samples exposed at five different temperatures (240, 250, 260, 270, and 300 °C) for four different durations (15, 30, 45, and 60 min). The samples were cooled under ambient conditions and tested at room temperature. The nominal size of the samples was 120 mm( $L$ ) × 20 mm ( $W$ ) × 4 mm( $T$ ). It is expected that thermal damage is induced in a more or less uniform manner over the sample volume.

The resonance mode under consideration in this study ( $\pm 1800$  Hz) is the fundamental flexural mode of a beam. It has a stress concentration in the middle of the sample and displacement nodes at a distance of  $0.224L$  from both edges, with  $L$  as the length of the sample (120 mm) (Rayleigh 1896; Gérardin and Rixen 1994). In the experimental setup, the sample is hung up by two nylon wires at the node lines and is excited at a pure tone by a loudspeaker (diameter of 32 mm the sound being concentrated by a converging cone of 180 mm length and 20 mm exit diameter) centered in the middle of the sample. The response is measured by a laser vibrometer (Polytec OFV303, decoder VD02) near one of the edges. All equipment is computer controlled and operated through LabVIEW and GPIB. The acquisition of the signal is realized by a PXI-5122 DAQ-card.

In the NRS experiment, the samples were excited by a 1000 period burst excitation at a given amplitude and with a frequency close to the fundamental flexural resonance frequency. The recorded signal was 0.6 s (120,000 points at a sampling rate of 200 kHz) of the reverberation of the sample after the excitation was stopped. Figure 5b shows a typical response from the start of the excitation to the steady state together with the reverberation. To achieve a high accuracy in the recording of the reverberation signal, a variable vertical range acquisition procedure was implemented based on an automated feedback of the instantaneous amplitude response. In this procedure, the dynamic range is decreased successively. At each



**Fig. 7** NRS results showing the analyzed frequency versus microstrain amplitude for the reference sample (showing almost no nonlinearity) and for three samples submitted on beforehand to different heating temperatures and exposure times: 250 °C for 30 min, 270 °C for 30 min, and 300 °C for 45 min. (Reproduced from Van Den Abeele et al. 2009, <https://doi.org/10.1121/1.3184583> with the permission of the Acoustical Society of America.)

range, signals are acquired and averaged ten times. The various signals recorded at decreasing dynamic range are finally matched to create a composed signal with adequate vertical resolution over the entire time axis. The amplitude, frequency, and damping information contained in the resulting signal are then analyzed by dividing the composed signal into several windows (with fixed time duration of 10 ms, which is typically of the order of 20 periodic oscillations) and by fitting the previously described exponentially decaying sine function (Eq. 7) to the data using a Levenberg-Marquardt algorithm to determine the parameters  $f_k$ ,  $\xi_k$ ,  $\phi_k$  and  $A_k$ , with  $k$  referring to the  $k$ th time window. This yields the evolution of the frequency  $f_k$  and damping characteristic ( $\xi_k$ ) as a function of the amplitude  $A_k$  in the decaying signal.

Figure 7 shows the results for the instantaneous resonance frequency versus amplitude for the reference sample (a), for two samples exposed for 30 min at 250 and 270 °C, respectively (b and c), and for a sample heated at 300 °C for 45 min (d). The analyzed data for the reference sample nearly follow a horizontal line, meaning that there is no or minimal dependence of the frequency on the amplitude. The reference sample is thus close to being a linear material. On the other hand, the results for longer exposure and higher temperature show an increased frequency dependence on amplitude, which indicates an increase in the material nonlinearity. Changing the window size for the analysis of the reverberating signal (within limits, of course) did not influence the results.

In order to quantify the degree of nonlinearity, the nonlinearity parameter  $\Gamma$  is defined as the proportionality coefficient between the relative resonance frequency shift and the strain amplitude  $\Delta\epsilon$ :

$$\frac{\Delta f}{f_0} = \Gamma \Delta\epsilon, \quad (8)$$

with  $f_0$  as the linear resonance frequency and  $\Delta f = f_0 - f_{res}$ . The strain amplitude values,  $\Delta\epsilon$ , were calculated from the measured particle velocity amplitude values,  $\nu$ , using the strain-velocity conversion expression for beams (Gérardin and Rixen 1994; Rayleigh 1896),

$$\Delta\epsilon \approx 0.219 \frac{T}{f\sqrt{12}} \left( \frac{4.73}{L} \right)^2 \nu \quad (9)$$

with thickness  $T = 4$  mm and length  $L = 120$  mm in the case of the used CFRP samples. It should be noted that because of the global character of the applied NEWS method,  $\Gamma$  only represents a global quantification of the nonlinearity, integrated over the whole sample. It contains no direct information on the localization of the defects. The values for the global NRS nonlinearity parameter  $\Gamma$  obtained in this study range from 0.6 to 6, and its variation as function of temperature and exposure time for all samples is summarized in Fig. 8a. We observed an overall increase with increasing exposure time and heat temperature up to a factor of 10 with respect to the reference value. The obtained values are comparable to values obtained for intact samples of heterogeneous materials such as slate (Van Den Abeele et al. 2000a), pultruded composites (Van Den Abeele et al. 2004a), concrete (Van Den Abeele and De Visscher 2000), and other materials (Johnson et al. 2004; Johnson and Sutin 2005).

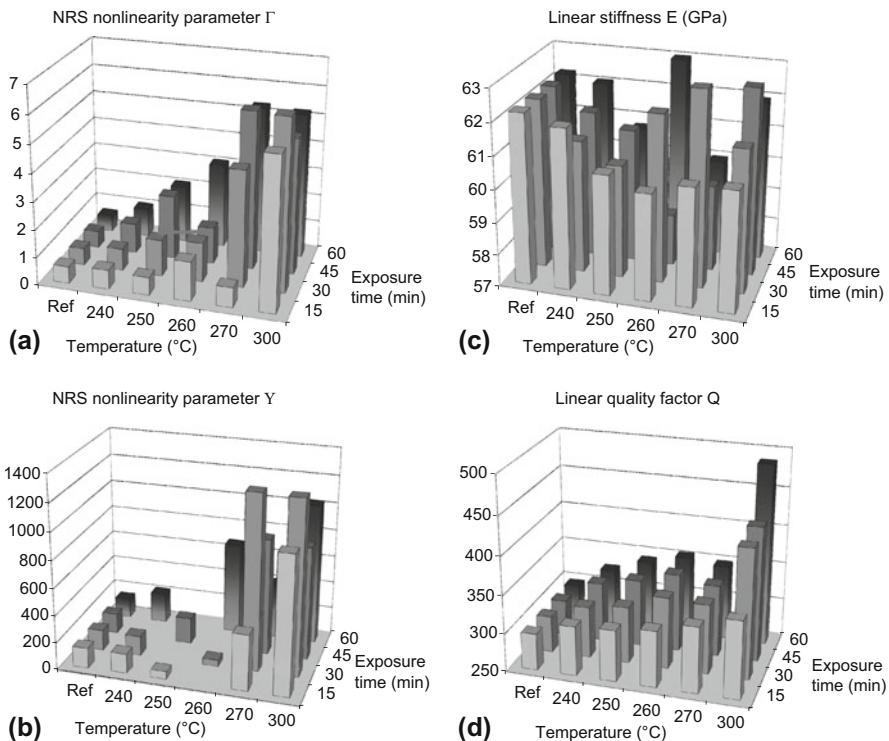
A similar behavior can be observed when analyzing the nonlinearity in the damping characteristic,

$$\frac{\Delta\xi}{\xi_0} = \Upsilon \Delta\epsilon \quad (10)$$

with  $\xi_0$  as the linear time constant (connected to the attenuation) leading to the damping nonlinear parameter  $\Upsilon$ . However, the errors in the analysis results are larger (support of the samples is very critical for attenuation), and the linear fits are not as clean as the ones dealing with the resonance frequency shift, which results in a less pronounced evolution (Fig. 8c).

## Comparison with the Linear Resonance and C-Scan Results and Discussion

The NRS analysis also provides the linear resonance signatures such as linear attenuation and linear resonance frequency. Ignoring subtle geometry changes, it is possible to calculate the global stiffness (Young's modulus  $E$ ) for the different samples from the linear resonance frequency values. However, a systematic



**Fig. 8** Summary of the NRS results for all 21 samples as function of heating temperature and exposure time: (a) NRS nonlinearity parameter deduced from the frequency response, (b) NRS nonlinearity parameter deduced from the damping response (some  $\Gamma$ -values with low repeatability are omitted), (c) linear values of the stiffness  $E$ , and (d) linear  $Q$ -factor (inverse attenuation). The reference point was duplicated for different exposure times to help visualize the trend of the evolution of the parameters with temperature. (Reproduced from Van Den Abeele et al. 2009, <https://doi.org/10.1121/1.3184583> with the permission of the Acoustical Society of America.)

change as function of the temperature and exposure time was not observed (Fig. 8b). For the attenuation, on the other hand, the linear value of the quality factor  $Q_0$  (inverse attenuation  $Q_0 = \pi f_0 / \xi_0$ ) increases with temperature and exposure time (Fig. 8d), meaning that the attenuation (at that frequency) decreases with increasing damage.

It is obvious that the nonlinear parameters derived in the NRS method show a considerable gain in sensitivity and provide a consistent interpretation of the results in contrast with the linear characteristics. The NRS results also show to be more sensitive to heat damage than traditional C-scan results as shown in Fig. 6. For 1 h exposures, a clear variation of the nonlinearity  $\Gamma$  is noticed for temperatures as low as 250 °C, whereas the C-scan data only gives decisive damage information from 285 °C.

## Quantification of the NRS Nonlinearity Parameter in Relation to the Microcrack Density

In order to compare the obtained values of the NRS nonlinearity parameter  $\Gamma$  with the microcrack density, five of the samples were sliced in the thickness direction and the crack density from each sample at the surface was calculated. To do so, the samples were imaged using light optical microscopy (LOM) coupled to a digital camera. Images are acquired with the magnification level set to 2. The process to quantify the crack density is described in detail in Van Den Abeele et al. (2009). In brief, the presence of cracks in layers with out-of-plane fiber orientation is determined analyzing gray-scale differences. In the selected areas, the crack-to-intact surface ratio, further referred to as crack density, is calculated.

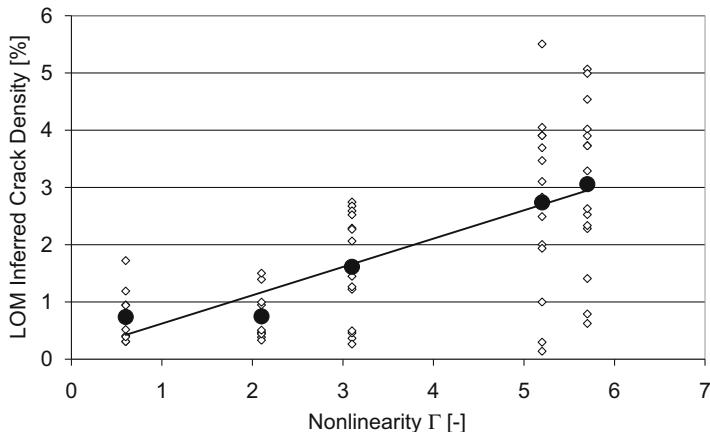
The results of this treatment applied to all images lead to Fig. 9. Open diamonds represent the different values of the crack density obtained in several subsets of the images. The spread of the results is mainly due to the small size of the subset area, which is analyzed and illustrates the statistical variation as function of the position along the surface. As one can expect, some areas show almost no cracks, while others exhibit several of them. The filled circles are the average value for each sample. A clear relationship between the NRS nonlinearity parameter  $\Gamma$  and the crack density can be observed.

1. The NRS nonlinearity parameter increases with increasing crack density.
2. The dispersion of the data increases with the NRS nonlinearity parameter. This can be explained by the nonhomogeneous repartition of the cracks inside the samples.
3. Even though the crack density measurements for the reference sample and the sample treated at 250 °C for 60 min are not significantly different, a vast increase (more than a factor of 2) in the NRS nonlinearity parameter is observed. This could imply that the crack density procedure based on the image treatment is not sensitive enough to identify the very early features (e.g., increase in dislocation nuclei) that are responsible for the increased NRS nonlinearity parameter, even though they definitely exist. It again illustrates the extreme sensitivity of nonlinear techniques to early stages of damage.

## Nonlinear Hysteretic Model

As mentioned in section “[Nonlinear Elasticity and Nondestructive Testing](#),” nonlinearity can be included by allowing the Young’s modulus to depend on the stress and stress rates (and if necessary other history-dependent variables). In the following simulations, the PM approach, described in section “[Nonclassical Nonlinearity: Hysteresis and the Preisach-Mayergoyz Approach](#),” is used to take the nonlinearity into account. The free vibration of the first eigenmode of an Euler-Bernoulli beam was implemented using a finite difference Scheme. A (linear) constant damping coefficient was introduced using a relaxation time mechanism described in Blanch et al. (1995).

Upon performing the numerical simulations, the exact size and mass were measured for each sample, yielding input values for  $\rho$  and  $I$ . Five relaxation



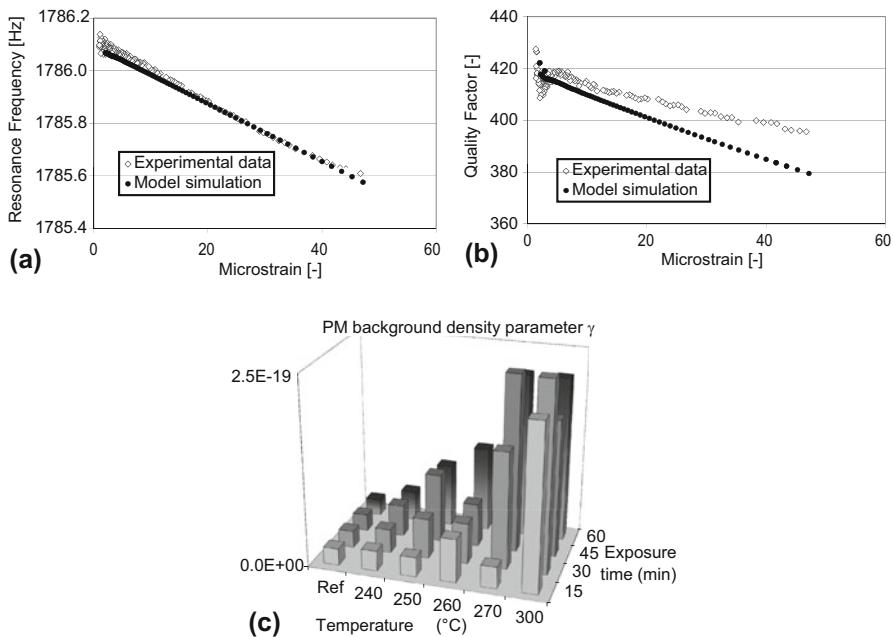
**Fig. 9** Crack density in selected areas vs. NRS nonlinearity parameter  $\Gamma$ . Open diamonds are values of the crack density obtained for different parts of the images. Filled circles are the averages for the whole surface for each sample. The line is the linear trend line for the average values. The horizontal error bars for the experimentally obtained quantity of  $\Gamma$  are of a few percent for an individual sample. (Reproduced from Van Den Abeele et al. 2009, <https://doi.org/10.1121/1.3184583> with the permission of the Acoustical Society of America.)

mechanisms are assumed to provide a constant  $Q_0$  value over a broad frequency range (0.1–5000 Hz). Further, the linear value of the Young's modulus  $K_0$  and the linear quality factor  $Q_0$  are adjusted to obtain the correct low amplitude values for each sample. They are assumed to be uniform over the beam length. The nonlinearity is introduced by specifying the statistical distribution of the bistable PM elements. The simplest way, which is most commonly used for dynamic processes, is to assume a uniform distribution of the elements. In this case, only one parameter is needed. We call  $\hat{\gamma}$  the PM background density parameter defined as

$$\hat{\gamma} = \rho_{PM}\gamma, \quad (11)$$

in which  $\rho_{PM}$  is the constant PM space density and  $\gamma$  is the strain contribution of one hysteron. The parameter  $\hat{\gamma}$  is expressed in units of  $\text{Pa}^{-2}$ ) (McCall and Guyer 1994; Guyer et al. 1995; Van Den Abeele et al. 2004b; Vanaverbeke and Van Den Abeele 2006) and its value is assumed to be uniform over the length of the sample (simulating a uniform distribution of damage). The physical meaning of the dimensionless quantity  $\hat{\gamma}d\sigma_c d\sigma_o$  is that it represents the deformation contribution of the hysteretic elements in the PM space with opening pressures between  $\sigma_o$  and  $\sigma_o + d\sigma_o$  and closing pressures between  $\sigma_c$  and  $\sigma_c + d\sigma_c$  upon switching from one state to the other (open to closed or closed to open). The larger  $\hat{\gamma}$ , the larger the nonlinear strain contribution. This is the only free parameter to be used for fitting the nonlinear behavior.

The comparison of the results for an exposure to 300 °C for 30 min is shown in Fig. 10. The simulations track the experimentally observed resonance frequency



**Fig. 10** Model results for CFRP\_300\_30 and comparison with experimental data: frequency (a) and quality factor (b). PM background density parameter  $\gamma$  used in the model simulations for all samples (c). The reference point was duplicated for different exposure times to help visualize the trend of the evolution of the parameters with temperature. (Reproduced from Van Den Abeele et al. 2009, <https://doi.org/10.1121/1.3184583> with the permission of the Acoustical Society of America.)

reduction extremely well. For the nonlinearity in the damping, the experimental data are generally noisier. Nevertheless, it shows more or less the same tendency as was predicted in the simulations (with a 5% error at 50 microstrain). The discrepancy could be due to the nonideal support of the beams by the nylon wires located at the node lines in the experimental setup.

Three important issues about the nonlinearity parameter quantification should be noted.

1. The use of classical reversible nonlinear models, such as the polynomial expansion of stress versus strain (or vice versa) (Guyer et al. 1998), would lead to a quadratic behavior of the resonance frequency shift with amplitude and does not affect the attenuation characteristic. To find the linear decrease observed in the data for the resonance frequency and the quality factor  $Q$ , it is essential to start from a hysteretic model.
2. The PM background density parameter  $\hat{\gamma}$  used in the numerical model is quite small. For the simulation of the nonlinear effects measured in the experiments, a value of  $\hat{\gamma}$  between  $2.2 \times 10^{-20} \text{ Pa}^{-2}$  (reference sample) and  $2.2 \times 10^{-19} \text{ Pa}^{-2}$  ( $300^{\circ}\text{C}$  for 60 min) was used. As mentioned above,  $d\sigma_c d\sigma_o$  represents the

deformation contribution of the hysteretic elements in the PM space with opening pressures between  $\sigma_o$  and  $\sigma_o + d\sigma_o$  and closing pressures between  $\sigma_c$  and  $\sigma_c + d\sigma_c$  upon switching from one state to the other. For a constant density in the statistical PM space, ranging from  $-5$  to  $5$  MPa, this would amount to a total hysteretic contribution to the strain of only

$$\hat{\gamma} \int_{-5 \text{ MPa}}^{5 \text{ MPa}} \int_{-5 \text{ MPa}}^{P_0} 1 d\sigma_c d\sigma_o = \hat{\gamma} \frac{1}{2} 10^{14} \quad (12)$$

$$\approx 10^{-6} - 10^{-5} \quad (13)$$

when changing the stress from  $-5$  to  $5$  MPa.

3. Based on the PM space approach (McCall and Guyer 1994; Guyer et al. 1995; Van Den Abeele et al. 2004b; Vanaverbeke and Van Den Abeele 2006), the relative modulus change is – at first order of approximation – proportional to the constant background density parameter  $\hat{\gamma}$  of the PM space, the linear modulus  $K_0$ , and the stress change itself. Since stress and strain are linked by the modulus, this results in

$$\frac{K_0 - K(\epsilon)}{K_0} \propto \hat{\gamma} K_0^2 \epsilon. \quad (14)$$

For those levels of nonlinear behavior observed in this study, giving rise to small frequency or modulus shifts, we indeed obtain in all cases a constant ratio between the macroscopically observed NRS nonlinearity parameter  $\Gamma$  and the theoretically found microscopic nonlinearity, which is expressed by the PM background density parameter  $\gamma$ :

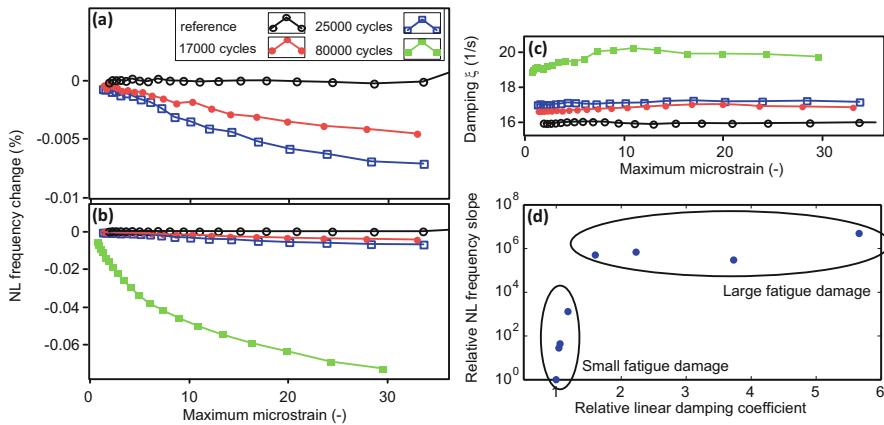
$$\frac{\Gamma}{\hat{\gamma}} \left( \frac{12}{K_0} \right)^2 = 1. \quad (15)$$

## Detection of Fatigue Damage in CFRP and Steel

### Fatigue Damage in CFRP Bars

Damage induced by thermal shocking is generally quite small and uniformly distributed over the samples. Damage induced by fatigue loading can be more severe and is usually highly localized. NRS can also be applied to detect and quantify fatigue damage, which is highly localized. A zone of delaminations and cracks in fatigued CFRP beams, and a single surface breaking crack in a steel landing gear bracket are investigated.

The study on the fatigued CFRP samples shows a clear and sensitive signature of the nonlinearity (reduction of the frequency with amplitude, and increasing



**Fig. 11** Nonlinear frequency shift for mechanically damaged CFRP samples (reference sample and three samples with increasing number of fatigue cycles) (a, b). Damping coefficient for the same samples (c). Relative increase of nonlinearity (slope of the frequency change) as a function of the linear attenuation signatures, showing the sensitivity of nonlinear signatures over linear signatures for small damage (d). (Reproduced from Van Damme and Van Den Abeele 2014, <https://doi.org/10.1007/s10921-014-0230-3> with the permission of the Springer Publishing Group.)

attenuation values) for different levels of damage (larger number of fatigue cycles), as can be seen in Fig. 11a, b. The frequency reduction is approximately linear in the case of a low number of fatigue cycles ( $< 25000$ ). However, to reach these nonlinear effects, the local level of microscopic nonlinearity  $\hat{\gamma}$  at the localized damage zone must be far larger than in the case of thermal damage. Confining the zone of nonlinearity to a region of length  $L/20$ , centered in the beam, acceptable agreement with experiments requires simulation input values for  $\hat{\gamma}$  of the order of  $2 \times 10^{-17}$  and higher, which is a factor 100 increase with respect to the thermally treated samples. When the number of fatigue cycles increases, secondary effects appear in the form of a reduced reduction rate and a stagnation of the frequency reduction (Fig. 11b). In the hysteretic PM space model this can be associated with a gradually decreasing population of the bistable elements in the statistical PM distribution, away from the PM space diagonal. This more complicated behavior cannot be described by a single parameter anymore. Similar effects are observed in the nonlinear attenuation.

The evolution of the linear damping parameter is consistent with the intuition that larger damage causes larger damping levels. The intercept at zero microstrain in Fig. 11c gradually provides a higher damping value for increasing fatigue cycles.

It is important to note, however, that the sensitivity of the nonlinearity signature is much larger than the sensitivity of a linear material property such as the attenuation, as can be appreciated from Fig. 11d. Additionally to the four previously described samples, four more fatigued samples were included in the analysis. Unfortunately, the fatigue history of these samples is unknown, but the damage is known to be much higher than the first series of four.

## Fatigue Damage in a Steel Steering Actuator Bracket

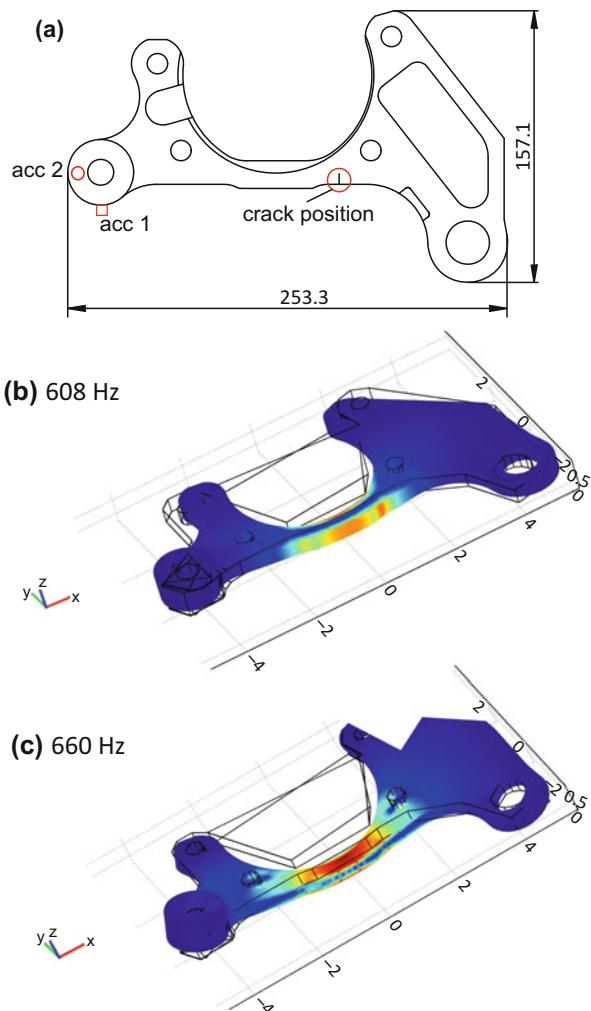
The final example illustrates the NRS technique on a real industrial sample with a complex geometry: a steering actuator bracket, which is part of the nose landing gear of a commuter airplane. The bracket is made of POLDI L-ROL low alloy Czech steel (Czech standard CSN 14331). The sample is represented in Fig. 12.

Three samples with a different damage degree are used:

**Bracket 1** was saved in intact state for comparison with brackets 2 and 3.

**Bracket 2** had accumulated 11,185 pulsating stress cycles (stress ratio  $R = 0$ ) using three point bending at a loading force level of 41,762 N. This level of the load caused a tensile stress about of 949 MPa in the critical point of the bracket according to strain gage measurement results. An estimation of crack growth

**Fig. 12** Steering actuator bracket used as an industrial sample for the comparison of NEWS techniques.  
(Reproduced from Van Damme and Van Den Abeele 2014, <https://doi.org/10.1007/s10921-014-0230-3> with the permission of the Springer Publishing Group.)



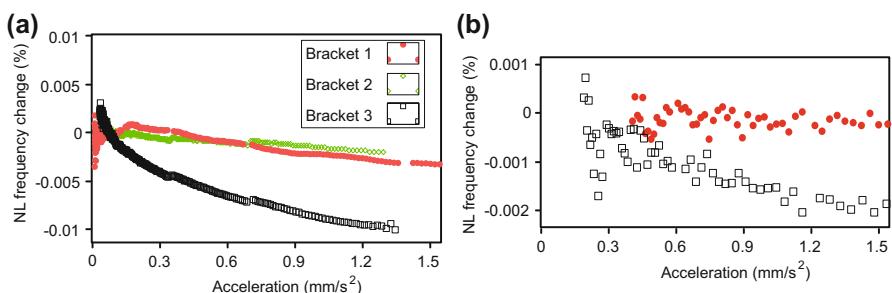
propagation was performed by calculations using the PREDIKCE code, VZLU in-house software (Aeronautical Research and Test Institute, Prague), and data obtained during fatigue loading of the bracket. The results of these calculations revealed inherent failure in the critical point of the bracket due to an initial crack. The estimated equivalent depth of the initial crack was about 0.7 mm.

**Bracket 3** was fatigued using three point bending until a visible crack was present. It needed in total 123,808 loading cycles of 43,762 N. This level of the load caused a tensile stress about of 870 MPa in the critical point of the bracket according to strain gage measurement results. The existence of a 2.1 mm fatigue crack was confirmed by an eddy current system.

The shape and dimensions of the samples and the position of the fatigue crack is shown in Fig. 12. The massive steel body of the test specimen is about  $12 \times 15$  mm in the zone of the crack location, which corresponds to the weakest point of the sample.

The experimental setup is essentially the same as for the CFRP samples except for the fact that the reception is made by means of an accelerometer (PCB Piezotronics 352B10). The results for the three samples are shown in Fig. 13. The experiment was performed twice, for two different eigenmodes of the bracket. Once again, there is a clear difference between the intact (red dots) and the damage bracket 1 (black squares). The measurement performed on the slightly damaged bracket 2 (green diamonds) is not decisive, as no significant difference with the undamaged bracket can be noticed. This is probably due to the fact that the single crack size is small compared to the applied wavelength, and the nonlinear wave distortion is too small to be measured. The scattering at low amplitude is due to the noise in the tail of the reverberation signal providing a bad resolution of the fits.

For this complex sample, with many vibration modes, the origin of the nonlinear wave distortion can be appreciated from the deformation shapes. The considered eigenmodes were modeled using finite element calculations. The von Mises stress



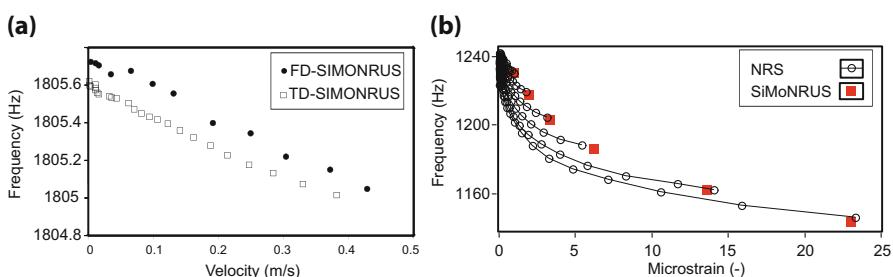
**Fig. 13** NRS analysis (instantaneous normalized resonance frequency as function of the response amplitude) for an intact (bracket 1) and two damaged (bracket 2 and 3) steering actuator brackets for two frequencies: 609 Hz (a) and 681 Hz (b). (Reproduced from Van Damme and Van Den Abeele 2014, <https://doi.org/10.1007/s10921-014-0230-3> with the permission of the Springer Publishing Group.)

for two modes is shown in Fig. 12. It is clearly visible that the first mode (607 Hz) is an in-plane bending mode, which allows the crack to open and close while vibrating. The second mode (660 Hz) is an out-of-plane torsional mode, which causes the crack's edges to slide over each other. It can be seen that the stress for both modes is large in the environment of the crack. Since the nonlinear frequency change of the first mode is much larger than the one of the second mode, one can assume that the first mechanism is more efficient than the second.

## Discussion

NRS can be considered a time-domain version of the well-established nonlinear SiMoNRUS method, since both use the amplitude dependent eigenfrequency shift in order to quantify material nonlinearity. Figure 14 indicates the consistency between the two resonance methods applied to the first bending mode of a CFRP beam. The analyzed resonance frequency from the sweeps exhibits the same slope as the fitted resonance frequency deduced from the reverberation signal. There is a small offset related to a slight change in experimental conditions. The advantage of the NRS method is that it requires fewer acquisitions (one time signal at a single excitation level versus a discrete frequency sweep at various increasing levels of excitation) and, by such, that it is faster than SiMoNRUS.

For all thermally damaged samples and all moderately fatigued CFRP samples (less than 25,000 cycles), the results of the SiMoNRUS and NRS techniques are consistent, independent of the initial excitation amplitude. However, once the fatigue level is high, we systematically observed a difference between the sweep-based SiMoNRUS results obtained at increasing levels of excitation and the resonance frequency analysis from the reverberating signals for similar initial levels of the excitation. Figure 14b shows a clear example of this observation for a highly damaged sample. The circles, connected with lines, correspond to the NRS analysis at different external excitation levels. The red squares correspond to the peak resonance frequency in the SiMoNRUS analysis at exactly the same levels of

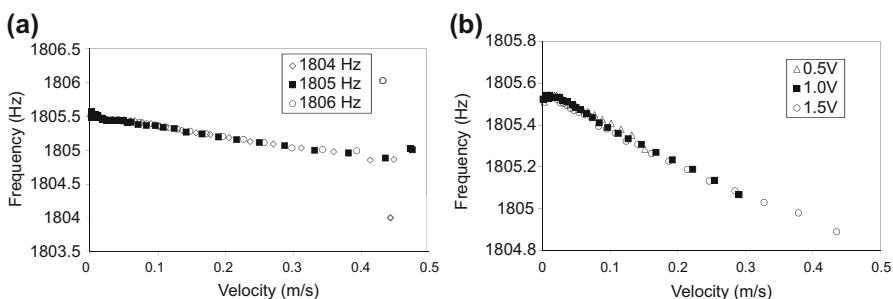


**Fig. 14** Illustration of the agreement between the SiMoNRUS (circles) and NRS (squares) results. Heat treated CFRP (300 °C for 30 min) (a) and highly fatigued CFRP (b). (Reproduced from Van Damme and Van Den Abeele 2014, <https://doi.org/10.1007/s10921-014-0230-3> with the permission of the Springer Publishing Group.)

external excitations. The highest amplitude data point of each NRS analysis matches quite well the results of the SiMoNRUS technique at the same excitation level.

The robustness of the method, a key feature for reproducible and trustworthy material evaluation, has been tested in several ways. Being a noncontact experiment, the only concern that could affect reproducibility is the string support of the sample. Paying particular attention to attach the supporting strings near the nodes of the resonance mode under consideration, several experiments were repeated after dismounting and remounting without invoking significant deviations in the results (errors of a few percent). In addition, even though experimental conditions may significantly affect the resonance frequency, the slope of the amplitude dependence, which yields the measure of the nonlinearity parameter in the NRS experiment, is independent of the exact resonance frequency value as it merely depends on the relative changes of it with respect to amplitude. These relative changes seem to be less dependent on the experimental conditions than the values of the resonance frequency. On top of this, we also verified that the obtained slope in the proportionality relations is independent of the chosen initial excitation frequency and applied voltage. The results of these investigations are illustrated in Fig. 15. In Fig. 15a, the response for three different frequencies close to the actual resonance frequency is illustrated for a fixed excitation amplitude. In Fig. 15b, the response at a fixed excitation frequency is illustrated for three different excitation amplitudes. This shows that the NRS nonlinearity parameter is independent of the initial excitation frequency (within limits in the order of the full width at half maximum of the resonance curve) and applied voltage (for regimes that do not involve slow dynamics). This insensitivity to changes in the experimental setup and conditions, in comparison to linear resonance measurements, is a practical advantage for the method.

In this chapter, the use of NRS was shown for two materials with very different properties, both regarding size, geometry, and material. Other researchers have further investigated the method, mainly to investigate the material nonlinearity due



**Fig. 15** Verification of the reliability of the NRS results for a single sample (CFRP\_300\_60) at various excitation frequencies (1804 Hz, 1805 Hz, 1806 Hz) for fixed amplitude (1.5 V) (a) and at various excitation amplitudes (0.5 V, 1 V, 1.5 V) for fixed excitation frequency (1805 Hz) (b). (Reproduced from Van Damme and Van Den Abeele 2014, <https://doi.org/10.1007/s10921-014-0230-3> with the permission of the Springer Publishing Group.)

to scattered microdamage. Successful measurements were performed on bone (Haupert et al. 2014; Muller and Renaud 2011). The evaluation of concrete poses additional experimental difficulties, since the samples are too large to excite in a noncontact way using a loudspeaker. In this case, repeated hammer strokes with increasing force were successfully used to excite the first bending mode (Leśnicki et al. 2011; Jin et al. 2017).

---

## Conclusion

In this chapter, a diagnostic technique for the detection of nonlinear material behavior was introduced and the efficiency of the NRS technique to detect damage was demonstrated. NRS is the time domain analogue of the SiMoNRUS technique, where the nonlinear amplitude dependent resonance frequency can be used to quantify material damage. Examples for global, heat induced, damage in CFRP beams were given, and a correlation was found between the nonlinear frequency changed and the optically measured crack density. This result confirms that the increase in nonlinearity is linked to an increased network of cracks and that the nonlinear signature is sensitive to microscopic alterations. Numerical simulations of bending resonances, using a hysteretic nonlinear constitutive relation within the sample support the results and relate the macroscopic NRS nonlinearity parameter to the microscopic PM background density of hysteretic elements (section “[Non-linear Hysteretic Model](#)”).

The technique was furthermore used for the detection of fatigue damage, both in CFRP beams and in a steel industrial sample (section “[Detection of Fatigue Damage in CFRP and Steel](#)”). The experiments prove that even the presence of a single fatigue crack shows up in the NRS results, if the crack is not too small. Due to the low frequencies used, and hence the large wavelengths present in the sample, a small single crack might be overlooked by this technique. In the complex shaped sample, the influence of the vibration mode on the NRS results was examined, showing that the NRS efficiency depends on the considered mode.

The NRS technique has the advantage to be fast, it has few or no restrictions on the sample geometry and in most cases it can be implemented in a fully noncontact manner (due to the low frequency nature of the method). On the other hand, it necessitates free (or at least steady, amplitude-independent) boundary conditions and is only applicable for low attenuation materials.

---

## References

- Aleshin V, Van Den Abeele K (2005) Micro-potential model for stress-strain hysteresis of micro-cracked materials. *J Mech Phys Solids* 53(4):795–824
- Aleshin V, Van Den Abeele K (2007) Microcontact-based theory for acoustics in microdamaged materials. *J Mech Phys Solids* 55(2):366–390
- Aleshin V, Van Den Abeele K (2009) Preisach analysis of the hertz-mindlin system. *J Mech Phys Solids* 57(4):657–672

- Antonets V, Donskoy D, Sutin A (1986) Nonlinear vibro-diagnostics of flaws in multilayered structures. *Mech Compos Mater* 15(5):934–937
- Bentahar M, El Aqra H, El Guerjouma R, Griffa M, Scalerandi M (2006) Hysteretic elasticity in damaged concrete: quantitative analysis of slow and fast dynamics. *Phys Rev B* 73(1):14,116
- Blanch J, Robertsson J, Symes W (1995) Modeling of a constant-Q – methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique. *Geophysics* 60(1):176–184
- Buck O, Morris W, Richardson J (1978) Acoustic harmonic generation at unbonded interfaces and fatigue cracks. *Appl Phys Lett* 33(5):371–373
- Courtney C, Drinkwater B, Neild S, Wilcox P (2008) Factors affecting the ultrasonic intermodulation crack detection technique using bispectral analysis. *NDT & E Int* 41(3):223–234
- Darling T, TenCate J, Brown D, Clausen B, Vogel S (2004) Neutron diffraction study of the contribution of grain contacts to nonlinear stress-strain behavior. *Geophys Res Lett* 31:L16,604
- Gérardin M, Rixen D (1994) Mechanical vibrations: theory and application to structural dynamics, vol 25. Wiley, New York
- Glioza A, Griffa M, Scalerandi M (2006) Efficiency of time-reversed acoustics for nonlinear damage detection in solids. *J Acoust Soc Am* 120:2506
- Guyer R, Johnson P (1999) Nonlinear mesoscopic elasticity: evidence for a new class of materials. *Phys Today* 52:30–36
- Guyer RA, Johnson PA (2009) Nonlinear mesoscopic elasticity: the complex behaviour of rocks, soil, concrete. Wiley, New York
- Guyer R, McCall K, Boitnott G (1995) Hysteresis, discrete memory, and nonlinear wave propagation in rock: a new paradigm. *Phys Rev Lett* 74(17):3491–3494
- Guyer R, McCall K, Van Den Abeele K (1998) Slow elastic dynamics in a resonant bar of rock. *Geophys Res Lett* 25(10):1585–1588
- Hamilton M (1986) Fundamentals and applications of nonlinear acoustics, in nonlinear wave propagation in mechanics. The American Society of Mechanical Engineers, New York
- Haupert S, Guerard S, Peyrin F, Mitton D, Laugier P (2014) Non destructive characterization of cortical bone micro-damage by nonlinear resonant ultrasound spectroscopy. *PLoS One* 9(1):e83,599
- Hyllengren F (2001) Tech. Rep. No. TEK01-0022, C.S.M. Materialteknik, Linköping
- Jin J, Moreno MG, Riviere J, Shokouhi P (2017) Impact-based nonlinear acoustic testing for characterizing distributed damage in concrete. *J Nondestruct Eval* 36(3):51
- Johnson P, Sutin A (2005) Slow dynamics and anomalous nonlinear fast dynamics in diverse solids. *J Acoust Soc Am* 117:124
- Johnson P, Zinszner B, Rasolofosaon P (1996) Resonance and elastic nonlinear phenomena in rock. *J Geophys Res* 101(B5):11,553–11,564
- Johnson P, Zinszner B, Rasolofosaon P, Cohen-Tenoudji F, Van Den Abeele K (2004) Dynamic measurements of the nonlinear elastic parameter  $\alpha$  in rock under varying conditions. *J Geophys Res* 109(B2):B02,202
- Kawashima K, Murase M, Yamada R, Matsushima M, Uematsu M, Fujita F (2006) Nonlinear ultrasonic imaging of imperfectly bonded interfaces. *Ultrasonics* 44:e1329–e1333
- Kazakov V, Sutin A, Johnson P (2002) Sensitive imaging of an elastic nonlinear wave-scattering source in a solid. *Appl Phys Lett* 81:646
- Kim J, Jacobs L, Qu J (2011) Nonlinear ultrasonic techniques for nondestructive damage assessment in metallic materials. In: Fu-Kuo Chang (ed) 8th International workshop on structural health monitoring 2011: condition-based maintenance and intelligent structures. Department of Aeronautics and Astronautics, Stanford University, DEStech Publications, Incorporated
- Korotkov A, Sutin A (1994) Modulation of ultrasound by vibrations in metal constructions with cracks. *Acoust Lett* 18(4):59–62
- Landau LD, Lifshitz EM (1959) Theory of elasticity. Pergamon, Tarrytown
- Lawn D, Brian R (1998) Nonlinear stress-strain curves for solids containing closed cracks with friction. *J Mech Phys Solids* 46(1):85–113
- Lesnicki KJ, Kim JY, Kurtis KE, Jacobs LJ (2011) Characterization of asr damage in concrete using nonlinear impact resonance acoustic spectroscopy technique. *NDT & E Int* 44(8):721–727
- Mason W (1969) Internal friction mechanism that produces an attenuation in the earth's crust proportional to the frequency. *J Geophys Res* 74(20):4963–4966

- Matzkanin G (1999) Heat damage in graphite epoxy composites: degradation, measurement and detection. *J Nondestruct Test Ultrason* (Germany) 4(3). <https://www.ndt.net/article/v04n03/ntiac/ntiac.htm>
- Mavko G (1979) Frictional attenuation: an inherent amplitude dependence. *J Geophys Res* 84 (B9):4769–4775
- McCall K, Guyer R (1994) Equation of state and wave propagation in hysteretic nonlinear elastic materials. *J Geophys Res* 99(B12):23
- Mix PE (2005) Introduction to nondestructive testing: a training guide, 2nd edn. Wiley, Hoboken
- Morris W, Buck O, Inman R (1979) Acoustic harmonic generation due to fatigue damage in high-strength aluminum. *J Appl Phys* 50(11):6737–6741
- Muller M, Renaud G (2011) Nonlinear acoustics for non-invasive assessment of bone micro-damage. In: *Bone quantitative ultrasound*. Springer, Dordrecht, pp 381–408
- Nagy P (1998) Fatigue damage assessment by nonlinear ultrasonic materials characterization. *Ultrasonics* 36(1–5):375–381
- Nazarov V, Radostin A (2015) Nonlinear acoustic waves in micro-inhomogeneous solids. Wiley, Hoboken
- Pecorari C (2003) Nonlinear interaction of plane ultrasonic waves with an interface between rough surfaces in contact. *J Acoust Soc Am* 113:3065
- Pecorari C (2004) Adhesion and nonlinear scattering by rough surfaces in contact: beyond the phenomenology of the preisach-mayergoyz framework. *J Acoust Soc Am* 116:1938–1947
- Rayleigh B (1896) *The theory of sound*. Macmillan, London
- Schmerr LW (2016) Probability of detection and reliability. Springer International Publishing, Cham, pp 685–695
- Sharma M, Tutuncu A (1994) Grain contact adhesion hysteresis: a mechanism for attenuation of seismic waves. *Geophys Res Lett* 21(21):2323–2326
- Shkolnik I (1993) Nondestructive testing of concretes: new aspects. *Nondestruct Test Eval* 10(6):351–358
- Solodov I, Busse G (2007) Nonlinear air-coupled emission: the signature to reveal and image microdamage in solid materials. *Appl Phys Lett* 91(251):910
- Sutin A, Donskoy D (1998) Vibro-acoustic modulation nondestructive evaluation technique. In: *Proceedings of SPIE*, vol 3397. International Society for Optics and Photonics, p 226
- Sutin A, Delclos C, Lenclud M (1995) Investigations of the second harmonic generation due to cracks in large carbon electrodes. In: *Proceeding of the 2nd international symposium on acoustical and vibratory surveillance methods and diagnostic techniques*, Senlis, pp 725–735
- Sutin A, Johnson P, TenCate J (2003) Development of nonlinear time reversed acoustics (nltra) for applications to crack detection in solids. In: *Proceedings of the word congress of ultrasonics*, pp 7–10
- Taylor R (2011) Fiber composite aircraft-capability and safety. <http://www.atsb.gov.au/media/27758/ar2007021.pdf>
- Ten Cate J, Shankland T (1996) Slow dynamics in the nonlinear elastic response of Berea sandstone. *Geophys Res Lett* 23(21):3019–3022
- Van Damme B, Van Den Abeele K (2014) The application of nonlinear reverberation spectroscopy for the detection of localized fatigue damage. *J Nondestruct Eval* 33(2):263–268
- Van Den Abeele K, De Visscher J (2000) Damage assessment in reinforced concrete using spectral and temporal nonlinear vibration techniques. *Cem Concr Res* 30(9):1453–1464
- Van Den Abeele K, Johnson P, Guyer R, McCall K (1997) On the quasi-analytic treatment of hysteretic nonlinear response in elastic wave propagation. *J Acoust Soc Am* 101:1885–1898
- Van Den Abeele K, Carmeliet J, TenCate J, Johnson P (2000a) Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage, part II: single-mode nonlinear resonance acoustic spectroscopy. *Res Nondestruct Eval* 12(1):31–42
- Van Den Abeele K, Johnson P, Sutin A (2000b) Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage, part I: nonlinear wave modulation spectroscopy (NWMS). *Res Nondestruct Eval* 12(1):17–30

- Van Den Abeele K, Carmeliet J, Johnson P, Zinszner B (2002) Influence of water saturation on the nonlinear elastic mesoscopic response in earth materials and the implications to the mechanism of nonlinearity. *J Geophys Res* 107(6):2121
- Van Den Abeele K, Carmeliet J, Van De Velde K (2004a) Inferring the degradation of pultruded composites from dynamic nonlinear resonance measurements. *Polym Compos* 22(4):555–567
- Van Den Abeele K, Schubert F, Aleshin V, Windels F, Carmeliet J (2004b) Resonant bar simulations in media with localized damage. *Ultrasonics* 42(1–9):1017–1024
- Van Den Abeele K, Katkowski T, Wilkie-Chancellor N, Desadeler W (2006) Laboratory experiments using nonlinear elastic wave spectroscopy (NEWS): a precursor to health monitoring applications in aeronautics, cultural heritage, and civil engineering. In: *Universality of non-classical nonlinearity*, Springer, New York, pp 389–409
- Van Den Abeele K, Le Bas P, Van Damme B, Katkowski T (2009) Quantification of material nonlinearity in relation to microdamage density using nonlinear reverberation spectroscopy: experimental and theoretical study. *J Acoust Soc Am* 126:963
- Vanaverbeke S, Van Den Abeele K (2006) Multiscale approach for simulating nonlinear wave propagation in materials with localized microdamage. In: AIP conference proceedings, vol 838, p 91
- Walsh J (1966) Seismic wave attenuation in rock due to friction. *J Geophys Res* 71(10):2591–2599
- Zagrai A, Donskoy D, Chudnovsky A, Golovin E (2008) Micro-and macroscale damage detection using the nonlinear acoustic vibro-modulation technique. *Res Nondestruct Eval* 19(2):104–128
- Zhang J (2016) Defect detection, classification, and characterization using ultrasound. In: *Structural health monitoring (SHM) in aerospace structures*. Elsevier, Amsterdam, pp 307–323
- Zumpano G, Meo M (2007) A new nonlinear elastic time reversal acoustic method for the identification and localisation of stress corrosion cracking in welded plate-like structures—a simulation study. *Int J Solids Struct* 44(11–12):3666–3684

---

**Part II**

**Optical Techniques**



# Surface Brillouin Scattering

11

Arthur G. Every and J. Darrell Comins

## Contents

Introduction .....	328
The Mechanisms for Brillouin Scattering .....	329
Elasto-optic Scattering in Transparent Media: Kinematic Conditions .....	329
Surface Ripple-Mediated Brillouin Scattering from Opaque Solids:	
Kinematic Conditions .....	331
Surface Ripple Scattering Intensity .....	334
Calculation of the Surface Green's Function $G_{33}(k_{  }, \omega)$ .....	334
The Surface Green's Function $G_{33}(k_{  }, \omega, d)$ for a Coated Solid .....	335
SBS in Semi-opaque Solids .....	339
Measuring Method and Equipment Requirements .....	341
Lasers .....	341
The Multipass Tandem Fabry-Pérot Interferometer .....	341
Calibration and Errors of Measurement .....	342
High Temperature Surface Brillouin Scattering .....	343
Extraction of Materials Properties from Surface Brillouin Spectra .....	344
Relation Between Measured Velocities and Elastic Moduli .....	344
Determination of Elastic Moduli of Isotropic Solids from SBS Spectrum .....	344
Determination of Elastic Moduli of Single Crystals from SBS Spectra .....	345
Determination of Elastic Moduli of Thin Supported Films from SBS Spectra .....	346
Some NDE Applications of SBS .....	348
SBS Measurement of Bulk Material Properties: .....	349
Surface Phononic Crystals .....	352
Determination of the Elastic Constants of Thin Supported Layers .....	353
Cross-References .....	357
References .....	357

---

A. G. Every (✉)

School of Physics, University of the Witwatersrand, Johannesburg, South Africa  
e-mail: [Arthur.every@wits.ac.za](mailto:Arthur.every@wits.ac.za)

J. D. Comins

School of Physics, Materials Physics Research Institute and DST-NRF Centre of Excellence in Strong Materials, University of the Witwatersrand, Johannesburg, South Africa  
e-mail: [Darrell.comins@wits.ac.za](mailto:Darrell.comins@wits.ac.za)

**Abstract**

This chapter provides a broad introductory overview of surface Brillouin scattering, a noncontact optical technique for probing the surface dynamics of solids, and thereby obtaining information about their near-surface mechanical properties. It is of particular use in studying solids whose surfaces are modified in some way, such as by the presence of a thin surface coating or the effects of radiation damage. The setup used for doing the measurements and the equipment requirements are described. The sought-after information on the surface properties is extracted from the small fraction of the incident light which undergoes a frequency shift on being scattered. Multipass Fabry-Pérot interferometry is used to resolve this component. The mechanisms for the light scattering are explained and the methods used for data analysis and recovery of elastic constants from the measured data are set out. Finally, a number of applications of technological importance are reviewed.

---

**Introduction**

Surface Brillouin scattering (SBS) is a noncontact nondestructive technique for measuring the near-surface mechanical properties of solids and thin supported layers. It reveals these by optically probing acoustic excitations within a micron of the surface. It is applicable to opaque or semi-opaque solids such as metals and semiconductors, but can also be applied to transparent solids, most readily by depositing a thin reflective coating on the surface. Small samples are employed, and these can be investigated as a function of temperature and pressure, etc. SBS has found a range of NDE applications, from measuring the bulk elastic properties of superalloys and other hard materials as well as a variety of important semiconductor materials, to investigating thin films used as protective coatings and in microprocessors and micro-mechanical devices. It is an attractive NDE technique for evaluating the effects of near-surface damage caused by polishing or ion-bombardment, and for studying natural and chemically induced oxidation.

Brillouin light scattering (BLS) is a form of Raman scattering in which light of angular frequency  $\omega_i$  impinges on a medium and a small fraction of the light is scattered with a change in frequency brought about by thermal (or quantum) fluctuations of the medium. The distinction is that, as conventionally understood, in Raman scattering the operative modes of vibration, called optic modes, have neighboring atoms moving in quite different or even opposite directions. In Brillouin scattering, on the other hand, it is acoustic vibrations with wavelengths comparable to the wavelength of the incident light, and with neighboring atoms vibrating almost in phase with each other, which effect the scattering. This difference impacts on the instrumentation that is used in these two domains. In Raman scattering the frequency shifts are of the order of 10 THz, and so normal spectroscopic techniques can be used to resolve the frequency shifted light (the so-called in-elastically scattered light) from the frequency un-shifted (elastically scattered) light. In Brillouin scattering, on the other hand, the frequency shifts are of the order of 10 GHz, and multipass

Fabry-Pérot interferometry is required to resolve the in-elastically scattered light. Also the fraction of the light in-elastically scattered is much smaller in Brillouin scattering than in Raman scattering, and so longer data collection times are required. Brillouin scattering requires a fairly high polish to the surface(s) of the sample to minimize diffuse scattering from surface asperities.

For transparent solids, Brillouin scattering is dominated by elasto-optic scattering from bulk acoustic modes. On the other hand, for opaque or semi-opaque solids the scattering takes place at the surface itself or in the near-surface region that the light is able to penetrate, and is referred to as surface Brillouin scattering (SBS). Because of the relatively small scattering volume involved in SBS, the intensity of SBS is generally much smaller than Brillouin scattering in transparent solids. The main focus of this chapter is on SBS.

The aim of this chapter is to present a broad introductory overview of the subject of SBS that can be usefully read by someone with a basic knowledge of optics and ultrasonics. For an individual wishing to proceed further, there are several extensive reviews of surface Brillouin scattering including Sandercock (1982), Dil (1982), Nizzoli and Sandercock (1990), Mutti et al. (1995), Comins (2001), and Beghi et al. (2012). This chapter is laid out as follows. In section “[The Mechanisms for Brillouin Scattering](#)” the two mechanisms for Brillouin light scattering, namely elasto-optic and surface ripple scattering are described and the kinematic conditions for bulk and surface scattering established. Surface ripple scattering is explored at some length, and the relation of the scattering intensity to the surface elastodynamic Green’s function established. SBS from coated solids is discussed, and distinction made between the behavior of an acoustically slow material on a fast substrate and a fast material on a slow substrate. Section “[Measuring Method and Equipment Requirements](#)” describes the SBS technique and equipment requirements. Section “[Extraction of Materials Properties from Surface Brillouin Spectra](#)” delves into the inverse problem of extracting near-surface mechanical properties of a solid from SBS spectra, and some algorithms for handling measured data are discussed. Finally, in section “[Some NDE Applications of SBS](#)” a range of different NDE applications of technological importance in which SBS has been employed are reviewed, including the measurement of the temperature-dependent elastic constants of silicon and a nickel-based superalloy, investigations of surface phononic crystals, and the determination of the elastic constants of thin supported films of the slow-on-fast and fast-on-slow variety.

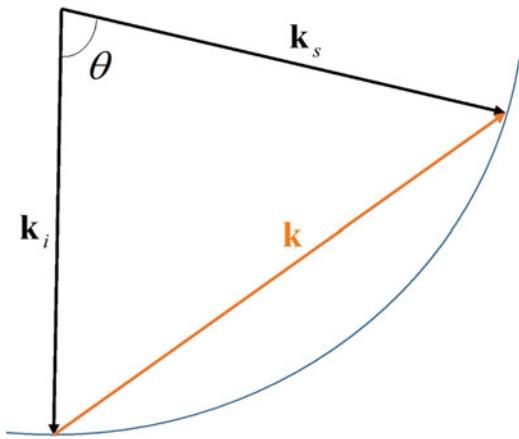
---

## **The Mechanisms for Brillouin Scattering**

### **Elasto-optic Scattering in Transparent Media: Kinematic Conditions**

While the main thrust of this article is on SBS, it is instructive to start with a brief discussion of Brillouin scattering in transparent media. Figure 1 depicts the scattering that takes place for a ray of light of angular frequency  $\omega_i$  and wave vector  $\mathbf{k}_i$  that is passing through a medium of refractive index  $n$ . The magnitude of  $\mathbf{k}_i$  is

**Fig. 1** Kinematic condition for Brillouin scattering, Eq. (3)



$$k_i = \frac{2\pi}{\lambda_i} = \frac{n\omega_i}{c}, \quad (1)$$

where  $\lambda_i$  is the wavelength and  $c$  is the velocity of light in vacuum.

At any finite temperature there are thermal fluctuations in all possible acoustic modes of vibration of the medium, accompanied by fluctuations in the elastic strain field  $\eta_{ij}$ , and through the elasto-optic effect, small fluctuations  $\delta\epsilon_{jk} = P_{jklm}\eta_{lm}$  in the dielectric constant of the medium, where  $P_{jklm}$  are the Pockels elasto-optic coefficients. In Brillouin scattering, the operative conditions are usually such that the acoustic vibrations which the light interacts with have frequencies  $f = \omega/2\pi \ll k_B T/h$ , where  $k_B = 1.38 \times 10^{-23} \text{ J/K}$  is Boltzmann's constant,  $T$  is the absolute temperature, and  $h = 6.63 \times 10^{-34} \text{ J.s}$  is Planck's constant. Under these conditions, classical statistical mechanics dictates that the average thermal energy per vibrational mode is  $k_B T$ , which is much greater than the vibrational energy quantum  $hf$ , and the scattering can be treated classically, which is what we will proceed to do. At very low temperatures (liquid He temperatures) or the much higher frequencies encountered in Raman scattering,  $hf \approx k_B T$ , and the scattering needs to be treated quantum mechanically. The quantum approach applied under the conditions of Brillouin scattering yields essentially the same result as the classical approach.

A fluctuation in the dielectric constant associated with an acoustic mode of wave vector  $\mathbf{k}$  and angular frequency

$$\omega = v k, \quad (2)$$

where  $v$  is the acoustic velocity, will result in a very small fraction, proportional to  $T$ , of the incident light being in-elastically scattered with wave vector

$$\mathbf{k}_s = \mathbf{k}_i \pm \mathbf{k}, \quad (3)$$

and angular frequency

$$\omega_s = \omega_i \pm \omega. \quad (4)$$

The vibrational frequency is much smaller than the optical frequency by a factor of order  $v/c \approx 10^{-5}$ , so  $\omega_i$  and  $\omega_s$  differ only very slightly, and  $\mathbf{k}_i$  and  $\mathbf{k}_s$  are to a good approximation equal in magnitude. So as shown in Fig. 1, which depicts the geometry of the scattering process, the tips of these two vectors lie on a circle, and are joined by the vector  $\mathbf{k}$ . The angle through which the light has been scattered has been labeled  $\theta$ . The scattering geometry is normally chosen by the investigator, so e.g., in the case of back-scattering  $\theta = \pi$ . It follows from Fig. 1 that

$$2k_i \sin(\theta/2) = k. \quad (5)$$

Combining this with Eqs. (1), (2) and (4) yields

$$v = \frac{\omega c}{2n\omega_i \sin(\theta/2)}. \quad (6)$$

Equation (6) allows one to calculate the acoustic velocity from the measured frequency shift. Generally there is also far more light which is elastically scattered at any angle due to inhomogeneities in the medium and other causes, and so what one actually observes at any scattering angle is a very intense spectral line at frequency  $\omega_i$  and then much less intense lines at frequencies  $\omega_i \pm \omega$ . The frequency down-shifted line is known as the Stokes component and the up-shifted line as the anti-Stokes component. There could be more than one line in each sideband if there is scattering from both longitudinal (L) and transverse (T) modes of a solid.

### Surface Ripple-Mediated Brillouin Scattering from Opaque Solids: Kinematic Conditions

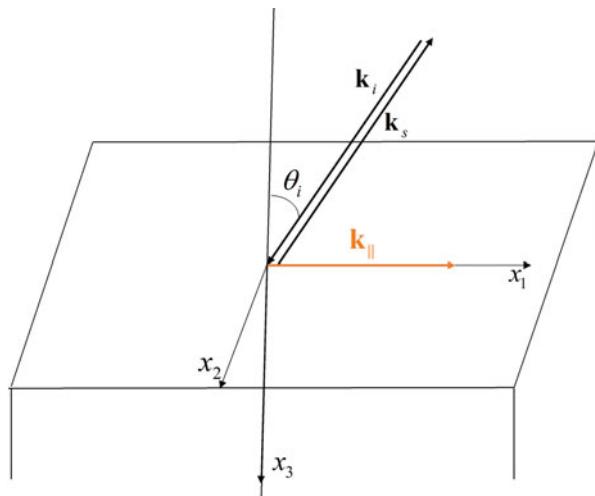
In the case of metals and other opaque solids, incident light penetrates only a very short distance into the medium and there is little elasto-optic scattering. What one observes instead is inelastic scattering from dynamic rippling of the surface derived from thermal fluctuations in Rayleigh and other guided surface modes and also from the continuum of bulk modes impinging on the surface. The spectrum of the scattered light conveys information about the near-surface mechanical properties of the solid, to a depth of order of the wavelength of the light. These may differ from the bulk properties because of surface treatment, oxidation, radiation damage, the presence of one or more thin over-layers of another substance or for other reasons. The usefulness of surface Brillouin scattering (SBS), however, extends also to the measurement of the bulk elastic properties of solids where, e.g., only small samples are available, provided that the near-surface properties don't differ significantly from the bulk properties. The theory of surface Brillouin scattering from opaque solids has

been developed by Loudon (1978a, b), Subbaswamy and Maradudin (1978), Velasco and Garcia-Moliner (1980), and Zhang et al. (1998a).

Figure 2 depicts the common back-scattering geometry that is employed in SBS measurements. A laser beam of frequency  $\omega_i$  and wave vector  $\mathbf{k}_i$  is incident on the highly polished surface of a sample at an angle  $\theta_i$  to the surface normal. Most of the light is specularly reflected without change in frequency. A small fraction of the light is in-elastically scattered in all directions by the dynamic rippling of the surface. Collection of scattered light, of frequency  $\omega_s$  and wave vector  $\mathbf{k}_s$  is carried out in the reverse direction. The spectrum of this scattered light generally consists of an intense central peak and then symmetrically disposed on either side of it much less intense Stokes and anti-Stokes sidebands. The central peak is comprised of diffusely scattered light of frequency  $\omega_i$ , and a slightly broadened component which is usually attributed to entropy fluctuations. The sidebands, which ideally are mirror images of each other about the central peak, display one or more sharp lines associated with Rayleigh, Sezawa and other guided surface modes, and a broad band known as the Lamb shoulder, given rise to by the continuum of bulk modes incident on the surface.

In SBS from an opaque solid, the scattering takes place at the surface and is caused by dynamic rippling of the surface, which can be thought of as comprised of a multitude of ripples of different wave vectors  $\mathbf{k}_{\parallel}$  parallel to the surface and frequencies  $\omega$ , which act as moving diffraction gratings. The incident light is thus scattered in all directions, and with Doppler changes in frequency determined by the ripples' phase velocities. Because the velocities of these acoustic ripples is 5 orders of magnitude smaller than the velocity of light,  $c$ , the change in frequency of the scattered light is proportionately small, and the same goes for the magnitude of the wave vector, and so to a good approximation  $k_s \approx k_i$ . It follows from the diffraction grating principle that the ripples responsible for back scattering

**Fig. 2** Backscattering geometry for SBS



have  $\mathbf{k}_{\parallel}$  in the plane of incidence of the light, and of magnitude the component of  $\mathbf{k}_s - \mathbf{k}_i$  in the surface, i.e.,

$$k_{\parallel} = 2k_i \sin \theta_i. \quad (7)$$

For a ripple travelling along the surface at the velocity  $v$ , its frequency is

$$\omega = v k_{\parallel}, \quad (8)$$

and this determines the change in frequency of the scattered light, i.e.,

$$\omega_s - \omega_i = \pm 2k_i \sin \theta_i v = \pm 2\omega_i \frac{v}{c} \sin \theta_i. \quad (9)$$

Guided surface modes such as Rayleigh and Sezawa waves have discrete velocities which are in most cases smaller than bulk wave velocities, and give rise to sharp lines in the SBS spectra at relatively small frequency shifts. Surface rippling is also caused by bulk L and T waves of velocities  $v_L$  and  $v_T$  respectively, being incident on and reflected from the surface at all angles. The phase velocity  $v$  of the surface ripple resulting from the incidence of a T wave at an angle  $\theta$  to the surface normal is

$$v = \frac{v_T}{\sin \theta}, \quad (10)$$

which ranges from  $v_T$  to  $\infty$ , depending on the angle. These ripples give rise to a broad band in the SBS spectrum, known as the Lamb shoulder, which for the anti-Stokes sideband extends from the T wave threshold at

$$\omega_s - \omega_i = 2\omega_i \frac{v_T}{c} \sin \theta_i, \quad (11)$$

to as far as the spectrum can be measured, with the Stokes sideband being the mirror image of this. The incidence of L waves adds a second broad band to the spectrum, extending from the L wave threshold at

$$\omega_s - \omega_i = 2\omega_i \frac{v_L}{c} \sin \theta_i, \quad (12)$$

to large frequencies. The abovementioned thresholds correspond to T and L waves respectively travelling parallel to the surface, which are known as transverse and longitudinal lateral waves. The situation for anisotropic solids is somewhat more subtle in that the lateral waves and their thresholds correspond to bulk waves with energy flux vectors (group velocities) parallel to the surface, which may in some cases have associated wave vectors oblique to the surface.

For ripple scattering the L threshold, because of mode conversion effects, shows up as a sharp dip in the Lamb shoulder as illustrated later in Fig. 4. It is useful, although not always possible, to be able to measure the position of this dip, since it yields the velocity of a bulk L wave travelling along the surface, the longitudinal

lateral wave or guided mode (LGM). This velocity, together with the Rayleigh velocity is sufficient to determine the two independent elastic constants of an isotropic solid. The T threshold tends to be a little more difficult to measure because of its proximity to the much more intense Rayleigh peak.

## Surface Ripple Scattering Intensity

It can be shown that in the classical regime the cross section for the scattering of light by the ripple mechanism with frequency change  $\omega$  and surface scattering wave vector  $\mathbf{k}_{\parallel}$  is proportional to the power spectrum  $\langle |u_3(\mathbf{k}_{\parallel}, \omega)|^2 \rangle$  of thermal fluctuations in the normal displacement of the surface profile (Loudon and Sandercock 1980). The fluctuation dissipation theorem is then commonly invoked to express the power spectrum in terms of the imaginary (dissipative) part of the Fourier ( $\mathbf{k}_{\parallel}, \omega$ ) domain surface elastodynamic response function (Green's function)  $\text{Im}G_{33}(\mathbf{k}_{\parallel}, \omega)$  for force and displacement normal to the surface. The scattering cross section is thereby given by

$$\frac{d^2\sigma}{d\Omega d\omega} = \frac{AT}{\omega} \text{Im}G_{33}(\mathbf{k}_{\parallel}, \omega), \quad (13)$$

where  $A$  is a constant that depends on the scattering geometry, frequency, and polarization of the incident light and optical properties of the medium, and  $T$  is the absolute temperature. Expressing the light scattering in terms of  $\text{Im}G_{33}(\mathbf{k}_{\parallel}, \omega)$  is simply a convenient computational strategy, and it should not be inferred from (13) that the incident light is causing the fluctuations. The light is simply probing existing thermal fluctuations.

## Calculation of the Surface Green's Function $G_{33}(\mathbf{k}_{\parallel}, \omega)$

SBS is often used to study elastically anisotropic solids such as crystals and textured polycrystalline films. It is therefore convenient to have at hand computer coding for calculating  $G_{33}(\mathbf{k}_{\parallel}, \omega)$  for anisotropic solids, which is equally applicable to an homogeneous isotropic solid as a special case. Comprehensive discussions of how  $G_{33}(\mathbf{k}_{\parallel}, \omega)$  is calculated, including the effects of an over layer, can be found in Comins (2001), Beghi et al. (2012), and Zhang et al. (1998a). In essence it is a matter of representing the response to a spatially and temporally periodic force acting on the surface as a linear superposition of three phase matched (i.e., sharing a common value of  $\mathbf{k}_{\parallel}$  and  $\omega$ ) plane wave solutions of the wave equation (Auld 1990)

$$\rho \frac{\partial^2 u_i}{\partial t^2} = c_{ijkl} \frac{\partial^2 u_k}{\partial x_j \partial x_l}, \quad (14)$$

where  $\rho$  is the density, and  $c_{ijkl}$  the elastic modulus tensor of the solid, subject to the boundary conditions.

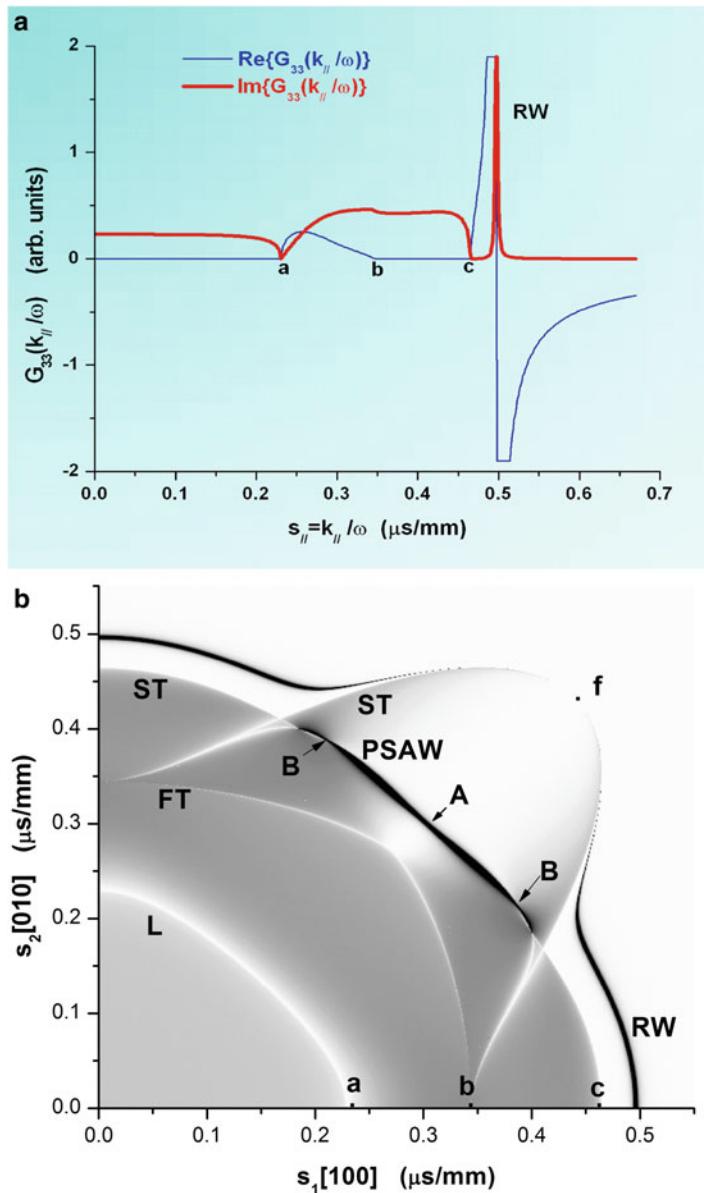
Figure 3a shows  $G_{33}(\mathbf{k}_{\parallel}/\omega)$  as a function of acoustic slowness  $s_{\parallel} = \mathbf{k}_{\parallel}/\omega$  at fixed frequency for the [100] direction in the (001) surface of the cubic crystal copper. The imaginary part  $\text{Im}G_{33}(\mathbf{k}_{\parallel}/\omega)$  has kinks at the limiting L (a), fast transverse FT (b) and slow transverse ST (c) slownesses (lateral waves), and is zero after the ST threshold (c) (transonic state), except at the Rayleigh wave slowness (RW) where it has a sharp spike associated with the pole of  $G_{33}(\mathbf{k}_{\parallel}/\omega)$ , which is determined by the vanishing of the boundary condition determinant (Farnell 1970; Farnell and Adler 1972). Some attenuation has been introduced to broaden the Rayleigh peak from a delta function and render it visible in the diagram.

Figure 3b shows  $\text{Im}G_{33}(\mathbf{k}_{\parallel}/\omega)$  rendered as a gray scale as a function of  $s_1 = k_1/\omega$  and  $s_2 = k_2/\omega$  at fixed frequency for the (001) surface of copper. Because of the fourfold crystallographic symmetry axis normal to the surface, the pattern repeats itself in the other three quadrants. Here we can see how the L, FT, and ST lateral waves vary with direction, undergoing splitting in some directions. The Rayleigh wave shows considerable variation with direction, both in slowness and in intensity. Toward the [110] direction the polarization of the RW tilts over toward the horizontal, and it becomes more weakly coupled to in  $G_{33}$  and hence decreases in intensity, ultimately vanishing as it degenerates with the limiting T bulk wave in the [110] direction at the point f. Also, as the RW approaches this degenerate condition, its displacement field penetrates ever more deeply into the solid, which also makes it more weakly coupled to in any component of the surface response. Extending out about 20° on either side of the [110] direction from point A is a sharp resonance in the bulk wave continuum, which pertains to a pseudo surface wave (PSAW). This is a wave consisting predominantly of two evanescent partial waves, but in addition there is a small bulk wave component which radiates the energy of this wave away from the surface, causing its attenuation. At the PSAW the boundary condition determinant is very small but non zero, except in the isolated directions A and B where the PSAW becomes a true supersonic surface wave (SSAW) (Farnell 1970).

Comparison of a measured and calculated spectrum for vanadium carbide ( $\text{VC}_{0.75}$ ) is shown in Fig. 4 (Zhang et al. 1998b). The calculated spectrum was conducted with optimized values of the elastic constants determined from the azimuthal angular dependence of the spectrum. The most prominent feature of this spectrum is the Rayleigh wave, and there is also a clear Lamb shoulder with a dip at the longitudinal guided mode (LGM) velocity. For many crystal orientations it is the PSAW that is observed rather than or in addition to the true SAW.

### The Surface Green's Function $G_{33}(\mathbf{k}_{\parallel}, \omega, d)$ for a Coated Solid

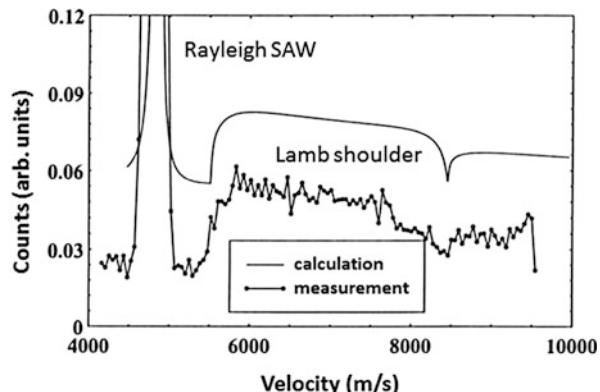
Considerable impetus was given to the study of the surface dynamics of coated solids by the review of Farnell and Adler (1972). One outcome is that SBS has become established as a valuable tool in the determination of the elastic properties of thin near-opaque sub-micron supported films that critically depend on the synthesis conditions. With the SBS frequency being in the range 1–50 GHz, and the wave field penetrating to about a wavelength below the surface, the SBS technique is ideal for studying such



**Fig. 3** (a)  $G_{33}(k_{\parallel}/\omega)$  for  $[100]$  direction in Cu(001) and (b) grey scale representation of  $\text{Im}G_{33}(k_{\parallel}/\omega)$  as a function of  $s_{\parallel} = k_{\parallel}/\omega$  for Cu(001)

films. The evaluation of  $\text{Im}G_{33}(k_{\parallel}, \omega, d)$ , used in the interpretation of SBS spectra for a supported film of thickness  $d$ , takes into account also the 6 phase matched partial waves in the film, and there are the 6 additional boundary conditions of continuity of displacement field and traction force at the interface, see Zhang et al. 1998a. The

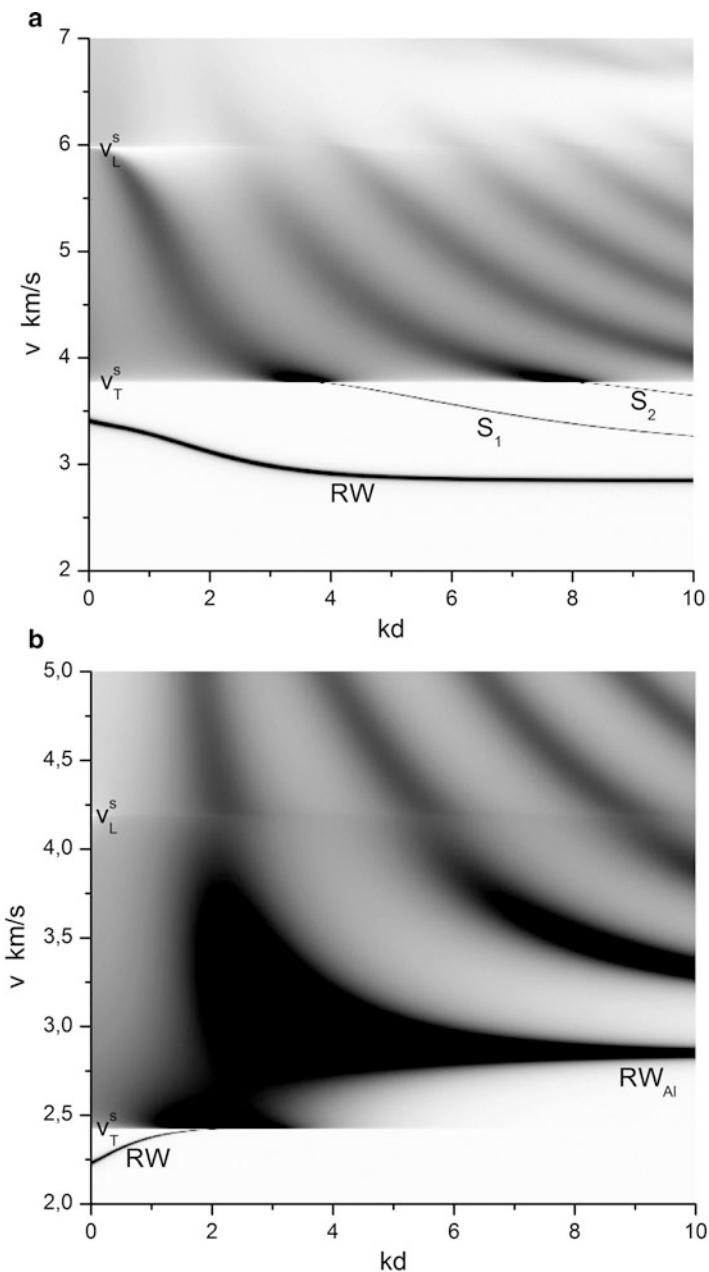
**Fig. 4** Measured and calculated SBS spectra of vanadium carbide ( $\text{VC}_{0.75}$ ) for  $k_{\parallel}$  in the  $[1\bar{1}0]$  direction in the (110) crystallographic surface. (From Zhang et al. 1998b)



velocity dispersion curves of the surface excitations are measured as a function of  $k_{\parallel}d$ , and appropriate fitting routines are employed to extract the set of elastic constants (see later). An open source program for calculating  $I = \text{Im } G_{33}(\mathbf{k}_{\parallel}, \omega, d)$  as a function of  $v = \omega/k_{\parallel}$  and  $k_{\parallel}d$  is available on request. The output is a  $480 \times 480$  array of integer values, allowing  $I$  to be depicted as a gray scale representation of the surface wave dispersion relation, as shown in Fig. 5a, b below, using ORIGIN or other suitable graphics software. Also generated are files of  $I$  versus  $v$  for a selection of values of  $k_{\parallel}d$ . The main program requires insertion of the values of the materials constants of the layer and substrate, the crystallographic orientation of the surface and scattering plane to be specified and the range of  $k_{\parallel}d$  and  $v$  to be provided.

It is well known from the review of Farnell and Adler (1972) that film-substrate combinations can broadly be classified into two types, slow-on-fast and fast-on-slow. The slow-on-fast category comprises combinations for which the bulk transverse velocity of the film is smaller than the transverse velocity of the substrate, which of course depends on the elastic moduli and densities of both components. The fast-on-slow category comprises combinations for which the bulk transverse velocity of the film is greater than the transverse velocity of the substrate.

As an example of a slow-on-fast combination, Fig. 5a depicts the dispersion relation for an isotropic aluminum layer on a fused quartz substrate. Below the T wave threshold of the substrate at  $V_T^s$ , the substrate partial waves for any  $k_{\parallel}d$  are both evanescent, and the modes that exist there, which are conditioned by the vanishing of the boundary condition determinant, are guided surface modes, with the displacement large in the layer, and falling off exponentially in the substrate. Starting at  $k_{\parallel}d = 0$ , which corresponds to the condition of a bare substrate, there is only one guided mode, the Rayleigh wave of the substrate. With increasing  $k_{\parallel}d$ , the Rayleigh wave falls off in velocity, ultimately for large  $k_{\parallel}d$  asymptotically approaching the Rayleigh velocity of the layer material. At a sequence of critical values of  $k_{\parallel}d$ , additional higher order guided modes, known as Sezawa modes and denoted  $S_1$  and  $S_2$  in Fig. 5a, emerge from the T wave threshold and fall off in velocity. The lowest of these asymptotically approaches the T velocity of the layer material for large  $k_{\parallel}d$ ,



**Fig. 5** Dispersion relations for (a) an aluminum layer on a fused quartz substrate and (b) an aluminum layer on zinc.  $v_T^s$  and  $v_L^s$  denote the substrate T and L threshold velocities respectively

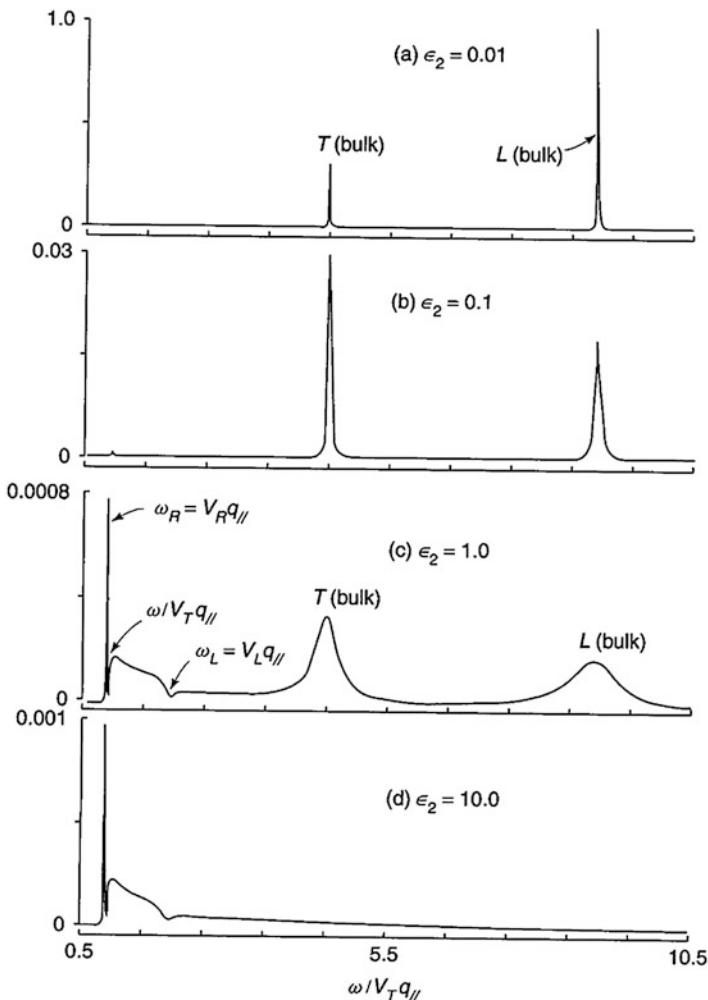
becoming the new T wave threshold. The existence of all these modes and their dependence on  $k_{\parallel}d$  allows for accurate determination of the properties of the layer. Above  $v_T^s$  the substrate T mode is a bulk mode and then above  $v_L^s$  the L mode of the substrate is also a bulk mode, allowing energy to be radiated away from the surface. Any guided modes that exist here tend to be leaky PSAW, damped to a lesser or greater extent. For certain combinations the PSAW are recognizable as extensions of the Sezawa modes into the radiative domain, and can be regarded as pseudo-Sezawa modes.

A different type of dispersive behavior is exhibited if the transverse velocity of the layer exceeds the transverse velocity of the substrate by a sufficient margin. As an example of a fast-on-slow combination, Fig. 5b depicts the dispersion relation for an aluminum layer on a zinc substrate. For this combination, starting at  $k_{\parallel}d = 0$  there is again only the one guided mode, which is the Rayleigh wave for the substrate. With increasing  $k_{\parallel}d$  the Rayleigh velocity trends upwards and eventually merges into the T bulk wave continuum of the substrate at a critical value  $k_{\parallel}d^{\text{crit}}$ . Beyond  $k_{\parallel}d^{\text{crit}}$  there are broadened PSAW to be observed above the T threshold but no guided modes below. For large  $k_{\parallel}d$  a progressively sharper feature evolves from the lowest PSAW, approaching asymptotically the Rayleigh velocity of the film material for large  $k_{\parallel}d$ . The situation is different if the layer is transparent and the substrate is opaque. In this case the ripple scattering takes place from the interface, and this feature may evolve into a Stoneley or pseudo-Stoneley interfacial mode for large  $k_{\parallel}d$ .

Not infrequently SBS is applied to transparent solids rendered opaque, or rather reflecting, by the deposition of a thin metallic film. A 30 nm layer of aluminum serves for this purpose, and doesn't represent excessive mass loading of the surface. Nevertheless for accurate determination of the properties of the substrate, the film has to be included in the calculations. A stack of two or three layers on a substrate is not uncommon, and in this case the evaluation of  $I = \text{Im } G_{33}(\mathbf{k}_{\parallel}, \omega, d_1, d_2, \dots)$  requires taking account of the 6 partial waves in each layer and the 6 additional boundary conditions at each interface (see Zizka et al. 2016).

## SBS in Semi-opaque Solids

In semi-opaque materials such as semiconductors, the presence of both bulk and surface Brillouin scattering can often be observed. The less opaque the material the greater the penetration of the light into the medium, and the greater the proportion of bulk elasto-optically scattered light in the spectrum. However, the fact that there is some attenuation of light in the medium means that the wave vectors of the incident and scattered light in the medium,  $\mathbf{k}_i$  and  $\mathbf{k}_s$ , are both complex. As a consequence the scattered light is broadened in frequency. Figure 6 shows combined surface and bulk Brillouin spectra for various values of the complex dielectric constant  $\epsilon = \epsilon_1 + i\epsilon_2$  of a solid calculated by Mills and Subbaswamy (1981). The attenuation is determined by the imaginary part of the dielectric constant  $\epsilon_2$ . Where this is very small, as shown in Fig. 6a, the bulk L and T scattering peaks are very sharp and there is little SBS. A slightly larger value of  $\epsilon_2$  yields slightly broadened bulk L and T peaks and the



**Fig. 6** Calculated combined surface and bulk Brillouin spectra for various values of the complex dielectric constant and  $\epsilon = \epsilon_1 + i\epsilon_2$ . (From Comins 2001, originally from Mills and Subbaswamy 1981)

trace of SBS Rayleigh peak. A still larger value of  $\epsilon_2$  yields significantly more broadened bulk L and T peaks and now a clear SBS spectrum, including the Lamb shoulder. Finally, a very large value of  $\epsilon_2$ , as shown in Fig. 6d yields a clear SBS spectrum and no trace of bulk scattering. From the point of view of determining materials properties, the intermediate situation of moderate  $\epsilon_2$  and hence absorption offers some advantage since the data it yields on both bulk and surface velocities provides more handles on the materials properties. It can also expose any differences there may be between bulk and near-surface properties.

## Measuring Method and Equipment Requirements

Surface Brillouin scattering measurements are conducted with monochromatic coherent laser light, and the spectral filtering is performed with multipass tandem Fabry-Pérot interferometry.

### Lasers

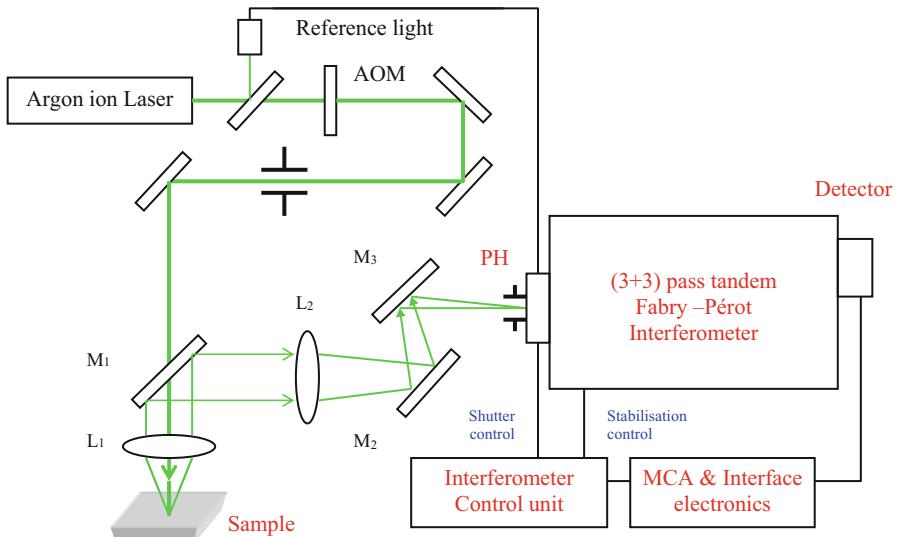
Argon-ion lasers are normally used and operated in the transverse electromagnetic single mode TEM<sub>00</sub> which gives a beam with a Gaussian intensity profile, minimum divergence, and the smallest spot diameter. To overcome mode hopping of the laser light, an intra-cavity etalon is used, this being a fixed-spacing Fabry-Pérot interferometer which allows the selection of a dominant frequency. In order to avoid the effects of temperature changes, which can result in instabilities of the laser frequency, the intra-cavity etalon is temperature controlled.

Recently compact and reliable Diode-Pumped Solid State (DPSS) lasers are being successfully used for SBS. An example is the Torus laser system of Laser Quantum emitting visible light at 532 nm with an output power up to 750 mW, being substantially lower than that of typical argon lasers, but usually adequate for SBS studies. The Torus tracks its single longitudinal mode position to avoid mode-hopping. In order to strongly attenuate the intensity of relatively spurious secondary laser modes, a temperature-stabilized etalon device (TFC) is used.

### The Multipass Tandem Fabry-Pérot Interferometer

The instrument currently used for Brillouin scattering is mostly the high contrast multipass tandem Fabry-Pérot (FP) interferometer developed by Sandercock. It is used in conjunction with a low-noise, high-sensitivity detector. The major advantage of the design of this instrument is the mounting of the tandem interferometers on a single scanning stage, which removes the difficulties of synchronizing the scans of separate interferometers. The reviews by Sandercock (1982), Mock and Guntherodt (1984), and Mutti et al. (1995) deal in considerable detail with the theory and operation of the (3 + 3) pass instrument. A (5 + 4) tandem interferometer with an even higher contrast is described by Lindsay et al. (1981). In a further development, a computer-controlled version of a Sandercock type of interferometer is discussed by Hillebrands (1999). This fully automated design provides control not only of the interferometer alignment but also the external optics, enabling unattended operation.

Figure 7 shows a block diagram of the Sandercock tandem FP and the collection optics used in our laboratory. The interferometer is dynamically isolated from the optical table on which the entire arrangement is mounted, through the use of an electronically controlled vibration isolation system. The scattered light is collected

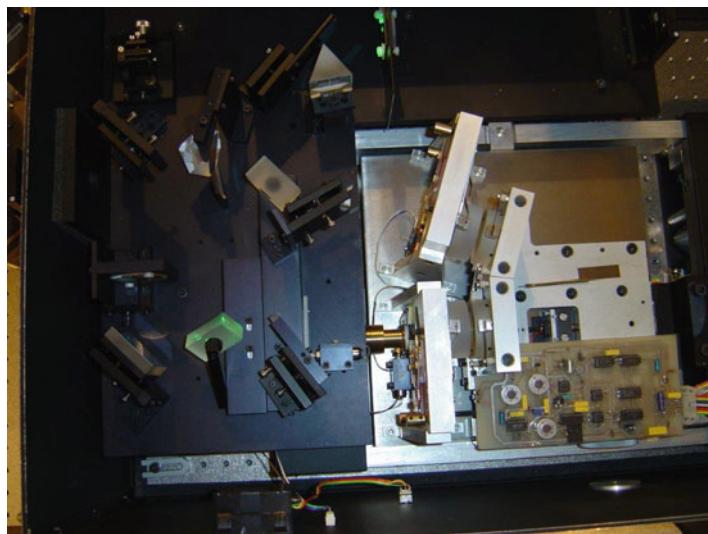


**Fig. 7** Block diagram of an experimental SBS arrangement

by an elliptical dielectric coated mirror  $M_1$  with a central 2-mm hole. This avoids spectral distortions caused by the use of a usual small steering mirror and its supporting arm, to direct the optical beam onto the sample, and it results in a gain in intensity. The auxiliary optics includes a shutter system with an adjustable window period in which the elastically scattered light is prevented from entering the detector during each scan. During the window the shutter system provides the control system signal necessary for automatic stabilization of the interferometer mirrors. The acousto-optic modulator allows the laser power on the sample to be continuously varied without affecting the control signal. The detector is a Laser Components Count Blue single photon counting module that offers a superior signal-to-noise ratio as opposed to the alternative of high-quality low-noise photomultiplier tubes (Fig. 8).

## Calibration and Errors of Measurement

A detailed discussion of the errors of measurement in surface Brillouin scattering is provided in the review by Mutti et al. (1995). They also discuss measurement accuracy in relation to geometrical aperture broadening which results from the use of the relatively large collection aperture. This aspect was extended by Stoddart et al. (1998) who comprehensively re-examined the effects of aperture broadening. They point out the importance of out-of-plane scattering and consider the dependence of the scattering cross section for surface ripple scattering on anisotropy of the surface. As their analysis shows, a limitation of the collection aperture significantly reduces



**Fig. 8** Aerial view of the JRS Instruments (3 + 3) multipass tandem FPI used in the experiments showing the tandem/scanning stage (white metal). The large arrow shows the position of the entrance pinhole for entry of the laser light. C marks the position of a translation stage for switching between alignment mode and tandem mode during the set-up procedure

the spectral distortions and resulting errors, but also reduces the light collection efficiency. The methods developed can thus provide guidance on the optimal sizes of collection aperture masks.

## High Temperature Surface Brillouin Scattering

Surface Brillouin scattering at high temperatures was pioneered by Stoddart et al. (1996) who measured the angular dispersion of the RW and the PSAW on the (001) plane of silicon as a function of temperature over the range 30–800 °C. The work was the first demonstration of SBS in high temperature measurements of elastic constants. A special high vacuum optical furnace was used which allowed the angular variation of the SAW velocity to be determined by the rotation of the sample around the surface normal. Subsequently there have been a number of measurements of this type.

As discussed by Faurie et al. (2017), there has been a need to extend surface Brillouin scattering to yet higher temperatures in order to determine the elastic moduli as a function of temperature. Such studies relate to high temperature stability (oxidation, spinodal decomposition of alloys, etc.) and their phase transitions (glass transition and crystallization). A Bühler high temperature chamber HDK S1 was specially modified for SBS measurements with a maximum temperature of 1600 K in a controlled atmosphere or high vacuum. The high temperature stage is operated by analyzing the backscattered light from the sample with a fixed angle of incidence.

## Extraction of Materials Properties from Surface Brillouin Spectra

### Relation Between Measured Velocities and Elastic Moduli

Brillouin scattering is applied to both isotropic and elastically anisotropic solids, particularly crystals and textured polycrystalline coatings. The objective of the measurements is usually to obtain the elastic moduli of the solid, and possibly how these depend on temperature and pressure, etc. This involves one in the solution of an inverse problem, the forward problem being the dependence of the surface wave velocities (Rayleigh, Sezawa and lateral waves, etc.) on the elastic moduli and densities of the components comprising the near-surface region. The forward problem is simplest for the surface of an homogeneous solid, and can be handled analytically for an isotropic solid and for high symmetry directions in an anisotropic solid.

### Determination of Elastic Moduli of Isotropic Solids from SBS Spectrum

In the case of an isotropic solid, there are two independent elastic moduli, which can be taken to be the Lamé constants  $\lambda$  and  $\mu$ , or the matrix elastic constants  $C_{11} = \lambda + 2\mu$  and  $C_{44} = \mu$  or any two of the engineering moduli: the shear modulus  $G = \mu$ , Poisson's ratio  $\nu = \frac{\lambda}{2(\lambda+\mu)}$ , Young's modulus  $E = \frac{\mu(3\lambda+2\mu)}{(\lambda+\mu)}$ , and bulk modulus  $K = \lambda + \frac{2}{3}\mu$  (Rose 1999; Fung 1965). So two independent measured quantities are in principle sufficient to determine all of these constants. The longitudinal guided mode (LGM) and transverse lateral wave velocities are given respectively by

$$v_L = \sqrt{\frac{C_{11}}{\rho}}; v_T = \sqrt{\frac{C_{44}}{\rho}}, \quad (15)$$

and their measurement and knowledge of the density  $\rho$  allow the elastic moduli to be determined. In practice though, while  $v_L$  can be determined from the LGM, i.e., the dip in the Lamb shoulder at the L wave threshold, the Rayleigh wave velocity  $v_R$  is far easier to measure accurately than  $v_T$ , because of the prominence of the Rayleigh wave peak in SBS spectra. In this case, with  $C_{11}$  determined by  $v_L$ , it remains to determine  $C_{44}$  from  $v_R$ . The Rayleigh velocity is conditioned by the vanishing of the Boundary condition determinant, the characteristic equation, which is a cubic equation for  $\xi = \rho v_R^2$ , that can be cast in the form (Every et al. 2010)

$$H(\xi) = \xi^2 C_{44}(C - \xi) - C_{11}(C_{44} - \xi)(C - C_{12}^2/C_{11} - \xi)^2 = 0, \quad (16)$$

where  $C_{12} = C_{11} - 2C_{44}$  and  $C = C_{11}$ . With  $v_R$  measured and thus  $\xi$  known, Eq. (16) is solved numerically for  $C_{44}$ . For positive Poisson's ratio (it cannot exceed 0.5), the ratio  $v_R/v_T$  lies in the range [0.874, 0.955], and so

$$1.096\xi < C_{44} < 1.309\xi, \quad (17)$$

limiting the required search range for  $C_{44}$ . If it happens that the LGM is difficult to measure, and only the Rayleigh velocity is available, Eq. (17) sets bounds on  $C_{44}$ , but  $C_{11}$  cannot be inferred with much confidence, unless suitable information from other sources is available. There are several approximate expressions for the dependence of the velocity ratio  $v_R/v_T$  on Poisson's ratio  $\nu$ , see e.g., Vinh and Malischewsky 2007. If one is to assume a typical value of 0.3 for  $\nu$ , the Viktorov approximation (Viktorov 1967)

$$\frac{v_R}{v_T} = \frac{0.87 + 1.12\nu}{1 + \nu}, \quad (18)$$

yields  $C_{44} = 1.162\xi$ . The value of  $C_{11}$  is then given by

$$C_{11} = \frac{2(1 - \nu)}{(1 - 2\nu)} C_{44}, \quad (19)$$

which is much more sensitive to  $\nu$  than  $C_{44}$  is.

## Determination of Elastic Moduli of Single Crystals from SBS Spectra

Cubic crystals have three independent elastic moduli  $C_{11}$ ,  $C_{12}$ , and  $C_{44}$ , and the other crystal classes a larger number. For a full determination of all the elastic constants of a crystal from SBS spectra, at least an equal number of independent measured quantities for various crystallographic planes and directions must be available. It is often the case that there are more measured quantities than elastic moduli, and so the elastic moduli are varied to obtain an optimized fit of calculated to measured data. For a least squares fit between measured and calculated velocities  $v_i^{\text{meas}}$  and  $v_i^{\text{calc}}$  respectively, the merit function takes the form

$$\chi^2 = \sum_i \frac{(v_i^{\text{meas}} - v_i^{\text{calc}})^2}{\sigma_i^2}, \quad (20)$$

where the  $\sigma_i$  are the standard deviations pertaining to the measured data. When using Rayleigh velocity data, Eq. (20) has the drawback that the Rayleigh velocities have to be calculated from the Rayleigh characteristic equation for each iteration of the elastic moduli in the optimization procedure.

Every et al. (2010) have proposed a somewhat simpler procedure, applicable to measurements made in high symmetry directions in orthorhombic and higher symmetry crystal classes, in which the merit function is taken as

$$\chi^2 = \sum_i w_i H_i^2(\xi_i). \quad (21)$$

The sum is over the characteristic functions determining the velocities, which are available in closed form in these directions, and the  $w_i$  are the relative weightings accorded to the individual terms. This obviates having to solve the characteristic equations for the Rayleigh velocities for insertion in (20). The characteristic functions for the Rayleigh velocities are given by (16) with  $C = C_{11}$  for the (100) direction and  $C = \frac{1}{2}(C_{11} + C_{12} + 2C_{44})$  for the (110) direction. The characteristic functions for the LGM are

$$H(\xi) = \xi - C, \quad (22)$$

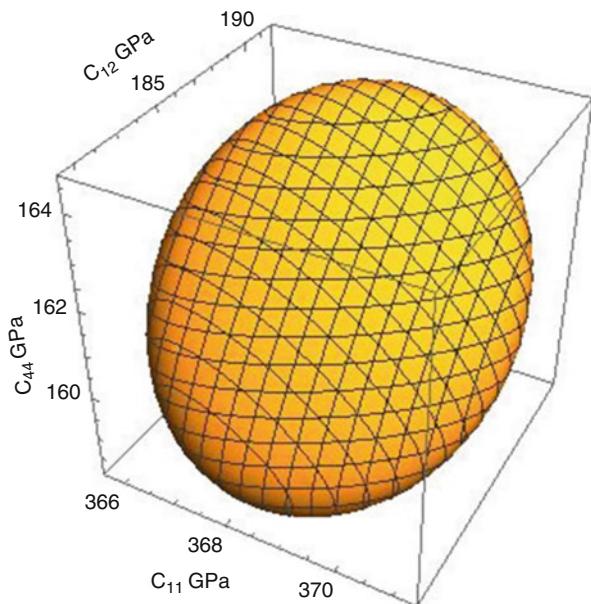
with the same  $C$ 's as for the Rayleigh wave.

Away from high symmetry directions in cubic and lower symmetry crystals there are no closed form expressions for the Rayleigh velocity, and the Green's function  $G_{33}(\mathbf{k}_{\parallel}, \omega)$  has to be evaluated numerically in the required directions for any given set of elastic moduli. These moduli are then varied to achieve a least squares fit between calculated and measured velocities. Having a greater number of measured velocities to fit should statistically improve the accuracy of the determination, but the actual fitting procedure can be cumbersome and time-consuming. Every et al. (2016) have developed a more methodical approach for determining elastic constants and their uncertainties, which they have illustrated using measured  $v_L$  and  $v_R$  data sets for the (001) surface of the cubic crystal  $\text{Rh}_3\text{Nb}$ . Existing closed form expressions for the LGM velocities in these symmetry planes, see e.g., Auld 1990, are used in the optimization. For the Rayleigh velocities, the Green's function approach, using good starting values of the elastic constants, is used to numerically calculate the velocities and their derivatives with respect the elastic constant variations for the directions of the measurements. These then allow a linearized approximation to be established for the Rayleigh velocities. With these closed form expressions for  $v_L$  and  $v_R$  a simple program, written in MATHEMATICA and using its optimization functions, is used to obtain the elastic constants and their uncertainties. Figure 9 shows the covariance ellipsoid in fitting the elastic constants of  $\text{Rh}_3\text{Nb}$ .

## Determination of Elastic Moduli of Thin Supported Films from SBS Spectra

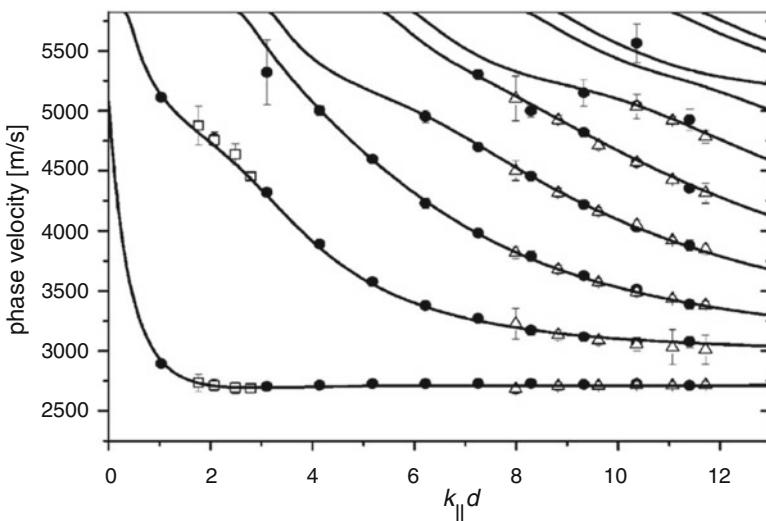
SBS has been widely applied to the determination of the elastic moduli, and in some cases also the density, of thin supported films. The better and more extensive the observed data, the more ambitious one can be in introducing free parameters. The simplest is to posit a single isotropic layer of known density on a known substrate, in which case there are just two free parameters,  $C_{11}$  and  $C_{44}$  of the film, to consider. Often textured polycrystalline films are

**Fig. 9** Covariance ellipsoid in the fitting of the elastic constants of  $\text{Rh}_3\text{Nb}$ . (From Every et al. 2016)



encountered which are transversely isotropic with symmetry axis normal to the surface. Although transverse isotropy is characterized by five independent elastic moduli, only four of them are accessible through SBS, which probes sagittal plane vibrations. These are  $C_{11}$ ,  $C_{13}$ ,  $C_{33}$ , and  $C_{55}$ . The density  $\rho$  is a possible fifth free parameter. SBS finds most success in its application to slow-on-fast combinations, with the film being opaque. Under favorable conditions one observes the  $k_{\parallel}d$  dependence of the Rayleigh wave and also one or more Sezawa's. With fast-on-slow combinations one is limited to the variation of the Rayleigh velocity before it merges into the substrate T threshold, and the appearance of pseudo-SAW resonances and their behavior for large  $k_{\parallel}d$ . The situation is further complicated for transparent films on an opaque substrate, in that interference effects between ripple scattering for the surface and interface and elasto-optic scattering in the film can tend to suppress some resonances. For a thick transparent film, scattering from an interfacial mode can be observed (Zinin et al. 1999).

An example of a successful characterization of a film using SBS, homogeneous tungsten carbide films were grown using RF sputtering with steps of differing thickness (60–655 nm) on a silicon substrate by Wittkowski et al. (2004, 2006). With the high mass density of WC, the system falls into the slow-on-fast category, yielding the Rayleigh SAW and many Sezawa modes. Figure 10 shows the measured data and calculated phase velocity dispersion curves that were obtained by adjusting



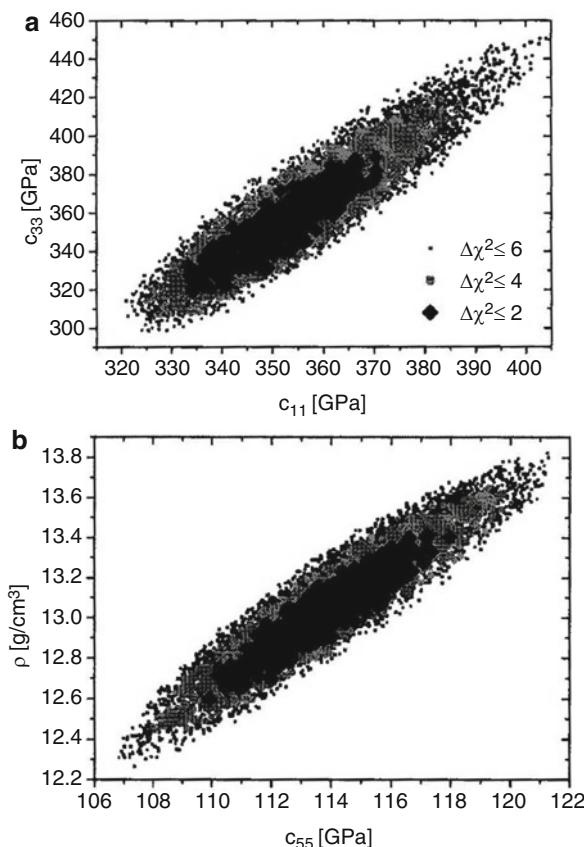
**Fig. 10** Phase velocity dispersion curves for tungsten carbide (WC) films on a silicon substrate, calculated using optimized values of elastic constants and density for the film. Measured data for various film thicknesses and scattering angles is shown by the symbols. The lowest branch pertains to the Rayleigh wave and the higher branches to various orders of Sezawa modes. (From Wittkowski et al. 2004)

the elastic constants and density of the film to obtain a best fit to the velocities, which yielded  $C_{11} = 349 \pm 7$  GPa,  $C_{13} = 126 \pm 8$  GPa,  $C_{33} = 348 \pm 13$  GPa,  $C_{55} = 114 \pm 2$  GPa, and density  $\rho = 13.0 \pm 0.2/\text{cm}^3$ . Figure 11 depicts the covariances in this fit by displaying projections on two parameter planes of parameter sets that yield values of  $\Delta\chi^2 = \chi^2 - \chi^2_{\min}$  that fall in different intervals. A conclusion that can be drawn from these is that SBS, though its measurement of acoustic velocities, is able to determine the ratios of elastic moduli to density more accurately than the moduli and density separately.

### Some NDE Applications of SBS

Surface Brillouin scattering is often applied to meet the needs of industry and scientific progress for information on the near-surface elastic properties of materials. It is a laboratory NDE technique that is applicable to small samples and allows measurement as a function of temperature and pressure, etc. It requires that the surface that is probed be highly polished to minimize the amount of diffuse scattering. It can be applied to the measurement also of bulk elastic properties where these do not differ significantly from the near-surface properties, or where the medium is semi-opaque, allowing bulk scattering also to be observed. A brief discussion of some of the applications to which SBS has been put follows.

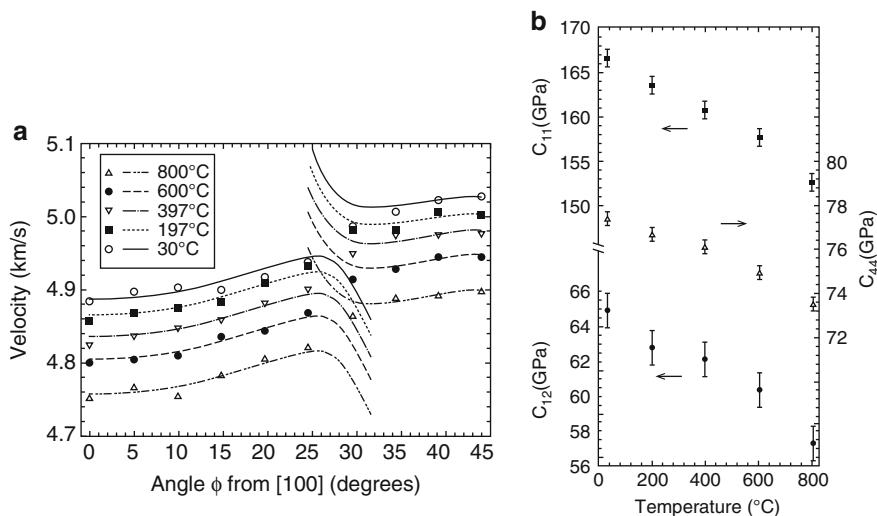
**Fig. 11** Covariances of the fit depicted by projections on two parameter planes of parameter sets that yield values of  $\Delta\chi^2 = \chi^2 - \chi^2_{\min}$  that fall in different intervals. (From Wittkowski et al. 2004)



## SBS Measurement of Bulk Material Properties:

### Crystalline and Polycrystalline Silicon at High Temperatures

The physical properties of this most important semiconductor material are certainly of importance. Stoddart et al. (1996) have used SBS to measure the angular dispersion of surface waves on Si(001) as a function of temperature over the range 30–800 °C as shown in Fig. 12a. As explained in section “[Calculation of the Surface Green’s Function](#),” for the cube surfaces of crystals such as Cu and Si, etc., the RW is prominent in an angular range around the [100] direction, while near the [110] direction, the PSAW dominates the scattering. There is a cross over region in between where sometimes both can be observed. This work was the first demonstration of SBS in the high temperature measurement of elastic constants, and used a specially designed optical furnace that allowed the angular variation of the SAW velocity to be determined. The three elastic constants of silicon  $C_{11}$ ,  $C_{12}$ , and  $C_{44}$  were determined at each temperature by least squares fitting to the measured data, and the results are shown in Fig. 12b. It is unusual to be able to obtain all three elastic



**Fig. 12** (a) Temperature dependence of the RW velocity (on the left) and PSAW velocity (on the right) on Si(001). The calculated curves are least-squares fits to the data. (b) Temperature dependence of the derived elastic constants of silicon. (From Stoddart et al. 1996)

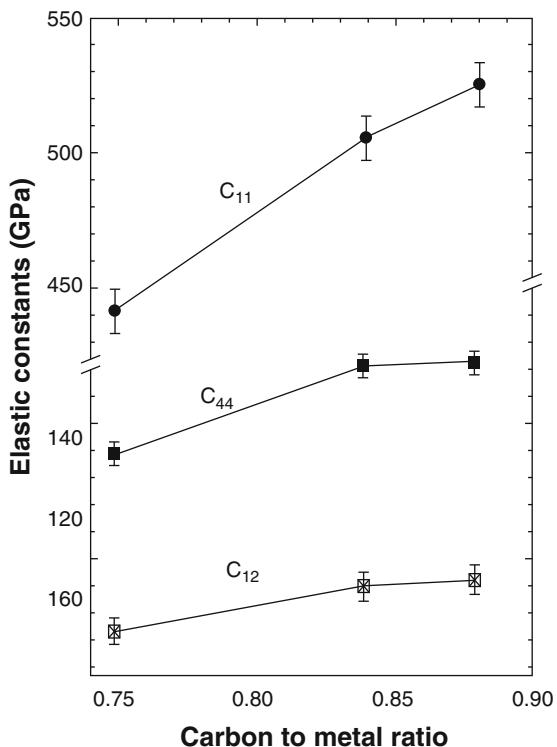
constants from SAW velocities alone, since these velocities are relatively insensitive to  $C_{11}$ . As a general rule having access the LGM velocity is required to ensure reliable values of  $C_{11}$  and the combination  $C_{12} + 2C_{44}$ .

Mathe et al. (2017) have used SBS to determine the two elastic constants and engineering moduli of a polycrystalline silicon (polysilicon) layer deposited on a Si(100) substrate and doped with 3000 ppm of boron, destined for micro-electromechanical systems (MEMS) applications. The layer thickness was about 50  $\mu$ , which is much greater than the SAW wavelengths probed, so the polysilicon could be regarded as bulk material. Their measurements of the RW and LGM velocities yielded the temperature dependence of the elastic constants in the range 20–110 °C.

### Vanadium Carbides

Vanadium carbides  $VC_x$  ( $0.65 < x < 0.90$ ), possess an unusual combination of physical and chemical properties, including high melting temperatures and strength, hardness, and good electrical and thermal conductivity. These properties promote their use in high temperature applications and for components requiring high wear resistance. They are cubic crystals conforming to the NaCl structure, with randomly distributed carbon vacancies which have a substantial effect on their elastic properties. Zhang et al. (1998b) have used SBS to determine the elastic constants of three different compositions of  $VC_x$ . Figure 4 shows their SBS spectrum for  $VC_{0.75}$  obtained with  $k_{\parallel}$  in the  $[1\bar{1}0]$  direction in the (110) crystallographic surface. They have extracted the elastic constants of their samples by least-squares fitting to the angular dependence of the Rayleigh and LGM velocities.

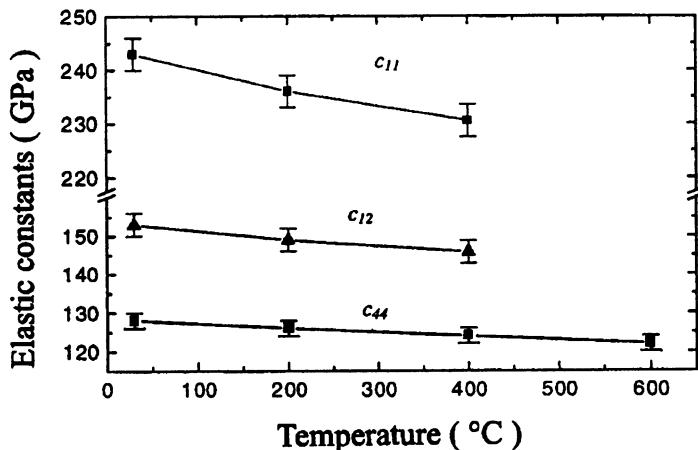
**Fig. 13** Variation of elastic constants with carbon-to-metal ratio in vanadium carbides. (From Zhang et al. 1998b)



Attempts to obtain the elastic constants from the RW velocity alone yielded a poorly defined minimum in  $\chi^2$ . Measurements were conducted on three compositions:  $VC_{0.88}$ ,  $VC_{0.84}$ , and  $VC_{0.75}$ . The elastic constants that were obtained are shown in Fig. 13.

### Single-Crystal Nickel-Based Superalloy CMSX-4

Nickel-based superalloys such as CMSX-4 are widely used for turbine blades in gas turbine engines in aircraft and for power generation. These blades retain their extraordinary strength and durability at high operating temperatures by being cast as single crystals. Zhang et al. (2001) carried out SBS measurements on a (001)-oriented single crystal of CMSX-4, which yielded the angular dependence of the Rayleigh SAW in a range near [100], PSAW in a range near [110], and the LGM velocity in the two principal directions. Using the SAW velocity alone they were not able to obtain the elastic constants with acceptable accuracy, but with the addition of the longitudinal velocity data they were able to obtain fits to the data with well defined minima to  $\chi^2$ . The analysis yielded ambient temperature elastic constants  $C_{11} = 243 \pm 2$  GPa,  $C_{12} = 153 \pm 2$  GPa, and  $C_{44} = 128 \pm 1$  GPa. Measurements were also carried out at higher temperatures, up to 600 °C, and yielded the results shown in Fig. 14.



**Fig. 14** Variation of the elastic constants of CMSX-4 with temperature. They show a linear decrease with temperature. (From Zhang et al. 2001)

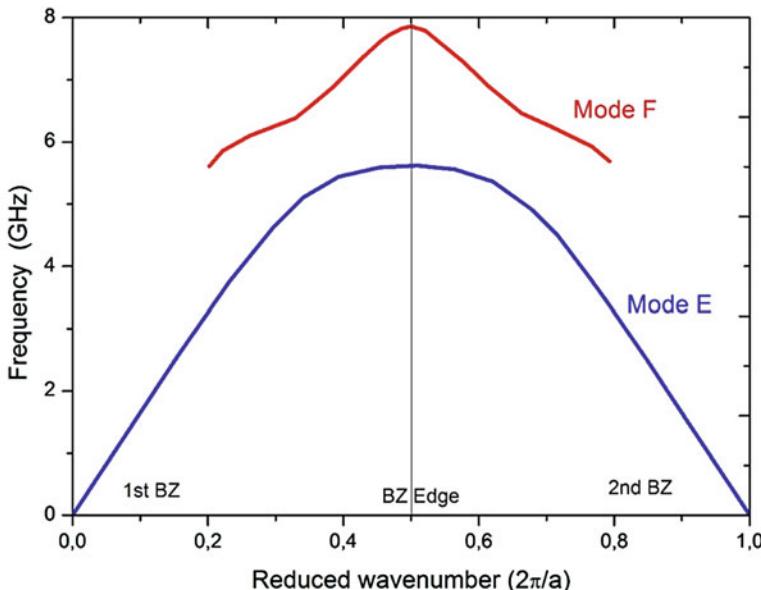
## Surface Phononic Crystals

An active area of application of SBS is in the study of surface phononic crystals (SPhC). Phononic crystals (PhC) are engineered structures which have one-, two- or three-dimensional periodicity in their mechanical properties. A key attribute they possess is that of frequency gaps in the transmission of acoustic waves. This has opened up areas of application in the construction of acoustic filters, transducers, and vibration shielding. PhCs offer the potential for acoustic cloaking and acoustic superlensing. SPhCs provide scope for similar control over surface acoustic waves. SPhCs encompass the modulation of the surface properties in 1D or 2D and can be constructed, for instance, by etching spatially periodic grooves in a surface or by depositing long strips or pillars of some material on a surface.

When a hitherto uniform surface is rendered periodic with repeat distance  $a$  in the  $x$ -direction say, then surface modes with wave vectors  $k + n\kappa$ ;  $n = 0, \pm 1, \pm 2, \pm 3, \dots$  for any given  $k$ , which are known as Bloch harmonics, are all coupled to each other by the non-uniformity. The quantity  $\kappa = 2\pi/a$  is called the reciprocal lattice vector. As a consequence of this, surface modes are distinguished by a wave vector  $k$ , not ranging between  $+$  and  $-$  infinity, but confined to the first Brillouin zone (BZ) located between  $-\kappa/2$  and  $+\kappa/2$ , and in addition a branch index  $N$  which runs through positive integers including zero. For a weak perturbation of an otherwise uniform surface, this is accomplished by ‘‘folding’’ the unperturbed dispersion relation  $\omega(k)$  back and forth across the BZ boundaries at  $\pm\kappa/2 = \pm\pi/a$ . The major effect of the periodic perturbation is initially to open up gaps in the dispersion relation at the zone boundaries.

## SBS of a 1D SPhC

A typical recent SBS investigation of an ideal 1D SPhC is that of Graczykowski et al. (2014). They have studied a grating in the form of rectangular grooves of depth 120 nm



**Fig. 15** Dispersion relation for SAW propagating in the  $[1\bar{1}0]$ -direction, perpendicular to the grooves of the SPhC. Schematic based on Fig. 5 of Graczykowski et al. (2014)

and width 60 nm made in a Si(001) surface. The grooves are aligned along the  $[110]$ -direction and the grating spacing is  $a = 300$  nm. With the SBS scattering vector  $k_{\parallel}$  aligned in the  $[1\bar{1}0]$ -direction parallel to the grooves they observe a number of modes whose dispersion relations they are able to fit with finite element calculations. There is no phase difference between the displacement fields in adjacent grooves, and because of the uniformity in the  $[110]$ -direction, there are no zone boundary effects.

With the scattering vector  $k_{\parallel}$  aligned in the  $[1\bar{1}0]$ -direction perpendicular to the grooves, because of the periodic variation of the surface in the  $[1\bar{1}0]$ -direction there is pronounced dispersion of the SAW near the Brillouin zone edge at  $\kappa/2$ , with the dispersion curves being horizontal at the boundary. The measurements have been done with a laser wavelength of  $\lambda_i = 532$  nm, with  $k_{\parallel} = 4\pi \sin \theta/\lambda_i = 0.0236 \sin \theta$  ranging over values both smaller and larger than  $\kappa/2$ , and so an extended Brillouin zone scheme encompassing the first and second BZs is needed for displaying and interpreting their results, as shown in Fig. 15. Their data is well accounted for with finite element calculations.

### Determination of the Elastic Constants of Thin Supported Layers

There have been numerous studies using SBS to determine the elastic constants and sometimes density or interfacial condition of supported solid layers. Most have concerned a single layer on a substrate, which falls either into the slow-on-fast category or the fast-on-slow category. There have also been some studies that have concerned

two and even three layers stacked on a substrate. Beyond this, there have been studies of superlattices of many alternating layers of different substances. In the following subsections brief consideration is given to a number of these investigations.

### **Low and High Dielectric Constant Materials**

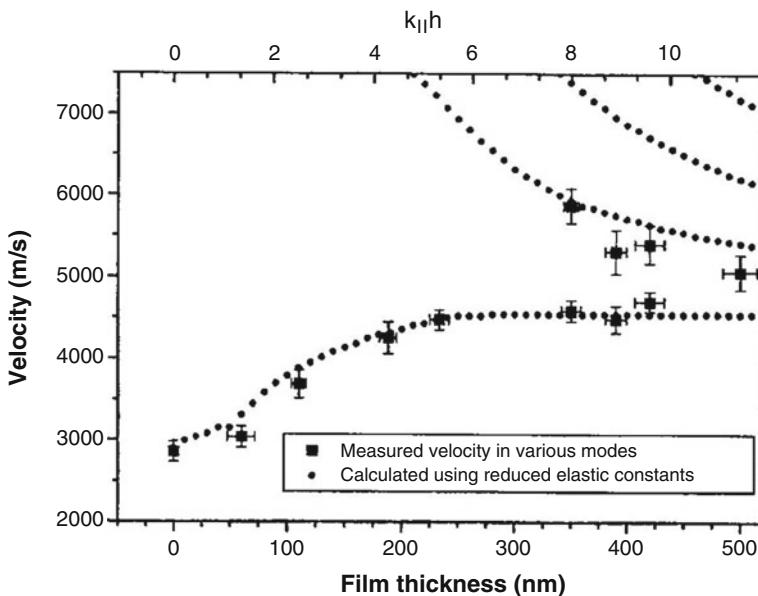
In the microelectronics industry new low and high dielectric constants ( $k$ ) materials in the form of thin films are finding important applications. Low- $k$  materials reduce interconnect capacitance and allow reduction in feature sizes, while high- $k$  dielectric materials allow for increased gate capacitance. Introducing the methyl group  $\text{CH}_3$  into  $\text{SiO}_2$  results in the porous  $\text{SiOC:H}$  structure and a halving of the value of  $k$ . But the mechanical strength of this material comes into question, particularly where TiN is used as a masking material for etching purposes. Similarly the mechanical properties of high- $k$  materials such as  $\text{HfO}_2$  and  $\text{BN:H}$  are of interest, since they influence device performance. SBS is able to provide information on these thin layers which is not skewed by challenges encountered in the more traditional indenter method. Zizka et al. (2016) have used SBS to investigate a number of these films, measuring the low lying Rayleigh wave and Sezawa modes, and for near normal incidence observing various L and T standing modes known as organ pipe resonances.

### **Superlattices**

Superlattices are periodically layered structures of alternating composition. They have numerous applications, including in semiconductors devices, thermoelectrics, in lasing, and have been used for generating coherent THz acoustic waves. There have been reports that certain superlattices have effective elastic constants exceeding the predictions of the effective medium theory of Grimsditch and Nizzoli (1986), which makes them interesting from a technological point of view. Continuum elasticity theory predicts that for acoustic wavelengths greatly exceeding the superlattice period  $\Lambda$ , measured velocities and inferred average elastic constants should be independent of  $\Lambda$ . However, there is evidence that for smaller values of  $\Lambda$  the interfaces play an important role in the physical properties, and deviations from continuum elasticity theory occur. A number of investigators have used SBS in an attempt to shed light on this issue (see De Bernabe et al. 2001 and references contained therein).

### **TiN on High-Speed Steel**

Titanium nitride is an extremely hard ceramic material that is used as coatings on cutting tools and other machine components, providing resistance to abrasion and corrosion. It is also used in the microelectronics industry for conductive connections and for diffusion barriers. It is non-toxic and finds applications in prosthetics. Pang et al. (1999) have conducted an extensive study of titanium nitride on high speed steel (TiN/HSS), for film thicknesses ranging from zero (bare substrate) to 4180 nm. This is a fast-on-slow combination which with increasing  $k_{\parallel}d$  starting from zero, sees the RW increasing in velocity and then merging with the bulk wave continuum of the substrate, thereafter becoming a PSAW. At higher

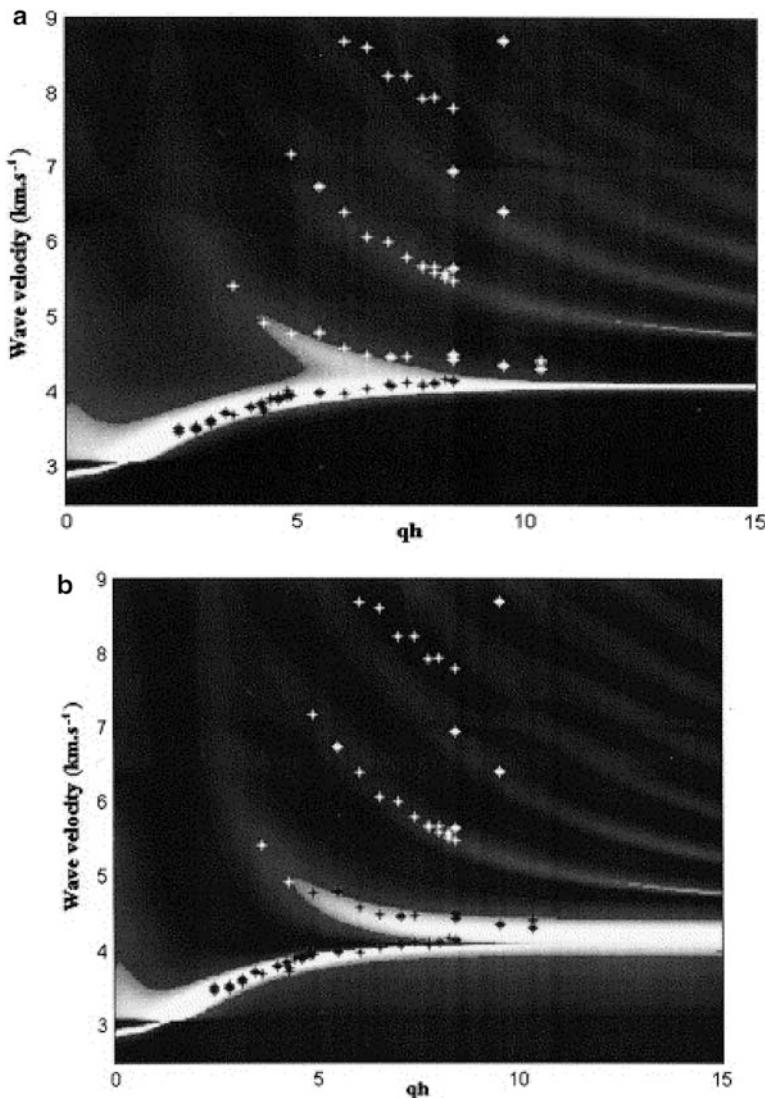


**Fig. 16** Comparison between measured and calculated wave velocities for TiN/HSS for film thicknesses less than 500 nm. (From Pang et al. 1999)

values of  $k_{\parallel}d$  a second PSAW appears. For large values of film thickness the lower PSAW tends to the Rayleigh velocity of bulk TiN. For layer thicknesses less than 500 nm the measurements indicate reduced effective values of the elastic constants for TiN. Figure 16 shows a comparison between measured and calculated wave velocities of TiN/HSS as a function of  $k_{\parallel}d$  for film thickness below 500 nm. The measured data is accounted for by assuming elastic constants for the film which are 75% of the values for bulk TiN, i.e., 75% of  $C_{11} = 475$  GPa and  $C_{44} = 179$  GPa.

### Barrier Film on Aluminum

Anodizing aluminum surfaces, which creates a thin layer of alumina on the Al surface, is widely used for electronic materials, packaging, lithography, adhesive bonding in aerospace applications, and elsewhere. Lefevre et al. (1999) have conducted a SBS study of barrier films on aluminum. Measurements were carried out on three samples for angles of incidence ranging from  $20^\circ$  to  $80^\circ$ . These films are transparent, and it is to be expected therefore that the dominant ripple scattering takes place at the interface. This is confirmed by their results, which are shown in Fig. 17. These films are acoustically faster than the substrate, and the measurements are confined to values of  $k_{\parallel}d$  (or  $qh$ ) beyond where the RW merges into the substrate cutoff. A number of PSAW are observed which are well accounted for by Green's function calculations. The lowest two PSAW are however dependent on whether the scattering is taken from the upper surface or interface. The observed results are better



**Fig. 17** SBS results for barrier film on aluminum. (a) Fitting of measurements to ripple scattering from surface of the film. (b) Fitting of measurements to ripple scattering from interface. (From Lefeuuvre et al. 1999)

accounted for by scattering from the interface, which at large  $k_{\parallel}d$  is associated with a highly damped pseudo-interfacial wave. Fitting to the data yielded the layer velocities  $v_T = 4595 \pm 25$  m/s and  $v_L = 7000 \pm 150$  m/s. Taking the density of the film to be  $3000$  kg/m<sup>3</sup> yielded a Young's modulus  $E = 142$  GPa and bulk modulus  $K = 63$  GPa.

The conditions for a true un-damped Stoneley interfacial wave, which of necessity is slower than the transverse velocities of both solids in contact, are fairly restrictive (Auld 1990). A Stoneley wave has been observed for molybdenum on fused silica by Bell et al. (1988), using SBS.

### Diamond-Like Carbon Films

Diamond-like carbon (DLC) exists in various forms, most importantly as tetrahedrally amorphous carbon (ta-C). It is widely used as a thin coating to impart favorable tribological features to the surfaces of metal components, such as increased hardness, resistance to abrasive wear, and reduced friction in sliding contact. There is a fairly extensive literature (see e.g., Ferrari et al. 1999; Manghnani et al. 2001; Berezina et al. 2004) on the use of SBS to investigate DLC synthesized under various conditions that admit different types of impurities and different ratios of tetrahedral carbon to graphitic carbon.

---

### Cross-References

- [Raman Scattering](#)
- 

### References

- Auld BA (1990) Acoustic fields and waves in solids. Robert E. Krieger Publishing Company, Malabar
- Beghi MG, Every AG, Prakapenka V, Zinin PV (2012) Chap. 10, Measurement of the elastic properties of solids by Brillouin spectroscopy. In: Kundu T (ed) Ultrasonic and electromagnetic NDE for structure and material characterization. CRC Press, Boca Raton, pp 539–610. ISBN: 978-1-4398-3663-7
- Bell JA, Zanoni R, Seaton CT, Stegeman GI, Macous J, Falco CM (1988) Elastic constants of, and Stoneley waves in, molybdenum films measured by Brillouin scattering. *Appl Phys Lett* 52:610
- Berezina S, Zinin PV, Schneider D, Fei D, Rebinsky DA (2004) Combining Brillouin spectroscopy and laser-SAW technique for elastic property characterization of thick DLC films. *Ultrasonics* 43:87–93
- Comins JD (2001) Surface Brillouin scattering. In: Levy M, Bass H, Stern R (eds) Handbook of elastic properties of solids, liquids, and gases, vol 1. Academic, New York, pp 349–378
- De Bernabe A, Prieto C, Caceres D, Vergara I, Every AG, Fischer HE (2001) Anomalous elastic properties of Si/Ge superlattices: the role of interfaces. *Phys Status Solidi (a)* 188:1023–1040
- Dil JG (1982) Brillouin scattering in condensed matter. *Rep Prog Phys* 45:285–334
- Every AG, Kotane LM, Comins JD (2010) Characteristic wave speeds in the surface Brillouin scattering measurement of elastic constants of crystals. *Phys Rev B* 81:224303
- Every AG, Sumanya C, Mathe BA, Zhang X, Comins JD (2016) Optimized determination of elastic constants of crystals and their uncertainties from surface Brillouin scattering. *Ultrasonics* 69:273–278
- Farnell GW (1970) Properties of elastic surface waves. In: Mason WP, Thurston RN (eds) Physical acoustics, vol 6. Academic, New York, pp 109–166
- Farnell GW, Adler EL (1972) Elastic wave propagation in thin layers. In: Mason WP, Thurston RN (eds) Physical acoustics, vol 9. Academic, New York, pp 35–127

- Faurie D, Girodon-Boulandet N, Kaladjian A, Challali F, Abadias G, Djemia P (2017) Setup for high-temperature surface Brillouin light scattering: application to opaque thin films and coatings. *Rev Sci Instrum* 88:023903
- Ferrari AC, Robertson J, Beghi MG, Bottani CE, Ferulano R, Pastorelli R (1999) Elastic constants of tetrahedral amorphous carbon films by Brillouin scattering. *Appl Phys Lett* 75:1893–1895
- Fung YC (1965) Foundations of solid mechanics. Prentice Hall, Eaglewood Cliffs
- Graczykowski B, Sledzinska M, Kehagias N, Alzina F, Reparaz JS, Sotomayor Torres CM (2014) Hypersonic phonon propagation in one-dimensional surface phononic crystal. *Appl Phys Lett* 104:123108
- Grimsditch M, Nizzoli F (1986) Effective elastic constants of superlattices of any symmetry. *Phys Rev B* 33:5891–5892
- Hillebrands B (1999) Progress in multipass tandem Fabry-Pérot interferometry: I. A fully automated, easy to use, self-aligning spectrometer with increased stability and flexibility. *Rev Sci Instrum* 70:1589–1598
- Lefevre O, Pang W, Zinin P, Comins JD, Every AG, Briggs GAD, Zeller BD, Thompson GE (1999) Determination of the elastic properties of a barrier film on aluminum by Brillouin spectroscopy. *Thin Solid Films* 350:53–58
- Lindsay SM, Anderson MW, Sandercock JR (1981) Construction and alignment of a high performance multipass Vernier tandem Fabry-Pérot interferometer. *Rev Sci Instrum* 52:1478–1486
- Loudon R (1978a) Theory of lineshapes for normal-incidence Brillouin scattering by acoustic phonons. *J Phys C Solid State Phys* 11:403–417
- Loudon R (1978b) Theory of surface-ripple Brillouin scattering by solids. *Phys Rev Lett* 40:581–583
- Loudon R, Sandercock JR (1980) Analysis of the light-scattering cross section for surface ripples on solids. *J Phys C* 13:2609–2622
- Manghnani MH, Tkachev S, Zinin PV, Zhang X, Brazhkin VV, Lyapin AG, Trojan IA (2001) Elastic properties of superhard amorphous carbon pressure-synthesized from  $C_{60}$  by surface Brillouin scattering. *Phys Rev B* 64:121403
- Mathe BA, Comins JD, Every AG, Hobbs LW (2017) High temperature surface Brillouin scattering study of mechanical properties of boron-doped epitaxial polysilicon. *AIP Adv* 7:025108
- Mills DL, Subbaswamy KR (1981) Surface and size effects on the light scattering spectra of solids. In: Wolf E (ed) Progress in optics, vol XIX. North Holland, Amsterdam, p 45
- Mock R, Guntherodt G (1984) Bulk elastic properties of metals determined by Brillouin scattering and its application to  $RCu_2Si_2$  (R=rare earth). *J Phys C Solid State Phys* 17:5635
- Mutti P, Bottani CE, Ghislotti G, Beghi M, Briggs GAD, Sandercock JR (1995) Surface Brillouin scattering – extending surface wave measurements to 20 GHz. In: Briggs GAD (ed) Advances in acoustic microscopy, vol 1. Plenum Press, New York, pp 249–300
- Nizzoli F, Sandercock JR (1990) Surface Brillouin scattering from phonons. In: Horton GK, Maradudin AA (eds) Dynamical properties of solids, vol 6. Elsevier, New York, pp 285–335
- Pang W, Every AG, Comins JD, Stoddart PR, Zhang X (1999) Brillouin scattering from acoustic excitations in TiN films on high speed steel – a stiffening system. *J Appl Phys* 86:311–317
- Rose JL (1999) Ultrasonic waves in solid media. Cambridge University Press, Cambridge, UK
- Sandercock JR (1982) Trends in Brillouin-scattering – studies of opaque materials, supported films, and central modes. In: Cardona M, Guntherodt G (eds) Light scattering in solids III. Recent results. Springer, Berlin, pp 173–206
- Stoddart PR, Comins JD, Every AG (1996) High-temperature studies of surface acoustic wave velocities in silicon by Brillouin scattering. *Physica* B219–220:717–719
- Stoddart PR, Crowhurst JC, Every AG, Comins JD (1998) Measurement precision in surface Brillouin scattering. *J Opt Soc Am B* 15:2481–2489
- Subbaswamy KR, Maradudin AA (1978) Photoelastic and surface-corrugation contribution to Brillouin scattering from an opaque crystal. *Phys Rev B* 18:4181–4199
- Velasco VR, Garcia-Moliner F (1980) Theory of surface waves in anisotropic cubic crystals. *J Phys C Solid State Phys* 13:2237–2256

- Viktorov IA (1967) Rayleigh and Lamb waves. Physical theory and applications. Plenum Press, New York
- Vinh PC, Malischewsky PG (2007) An improved approximation of Bergmann's form for the Rayleigh wave velocity. *Ultrasonics* 47:49–54
- Wittkowski T, Distler G, Jung K, Hillebrands B, Comins JD (2004) General methods for the determining of the stiffness tensor and mass density of thin films using Brillouin light scattering: study of tungsten carbide films. *Phys Rev B* 69:205401
- Wittkowski T, Jung K, Hillebrands B, Comins JD (2006) Structural and chemical phase transitions in tungsten carbide films evidenced by the analysis of their stiffness tensors. *J App Phys* 100:073513
- Zhang X, Comins JD, Every AG, Stoddart PR, Pang W, Derry TE (1998a) Surface Brillouin scattering study of the surface excitations in amorphous silicon layers produced by ion bombardment. *Phys Rev B* 58:13677–13685
- Zhang X, Comins JD, Every AG, Stoddart PR (1998b) Surface Brillouin scattering studies on vanadium carbide. *Int J Refract Met Hard Mater* 16:303–308
- Zhang X, Stoddart PR, Comins JD, Every AG (2001) High temperature elastic properties of a nickel-based superalloy studied by surface Brillouin scattering. *J Phys Condens Matter* 13:2281–2294
- Zinin P, Manghnani MH, Tkachev S, Askarpour V, Lefevre O, Every A (1999) Brillouin spectroscopy of surface in thin film  $\text{Si}_3\text{N}_4$  on GaAs. *Phys Rev B* 60:2844–2850
- Zizka J, King S, Every AG, Sooryakumar R (2016) Mechanical properties of low- and high- $k$  dielectric thin films: a surface Brillouin light scattering study. *J Appl Phys* 119:144102



# Interferometric Methods in NDE

12

Krzysztof Patorski and Maciej Trusiak

## Contents

Introduction .....	362
Encoding Information Using Two-Beam Interference .....	363
Interferometry with a Reference Beam .....	364
Interferometry with a Displaced Replica of Test Beam .....	365
Interferometry Using Conjugate Wave Front Beams .....	365
Information Decoding: Automatic Fringe Pattern Analysis .....	365
Time-Average Interference Microscopy for Microelement Vibration Testing .....	370
Moiré Interferometry for Experimental Mechanics and Material Engineering Evaluations .....	374
Summary .....	377
References .....	379

## Abstract

Optical interferometry offers unlimited research and testing possibilities for scientific and technological endeavors. Unprecedented progress in optoelectronics (new light sources and detectors), fiber optics, modern mechanical, and electronic hardware components as well as in computer engineering and informatics aids continuous development of automated, compact, full-field, non-contact, high-speed, and high-accuracy measurement systems. This chapter presents the theoretical description of basic two-beam interference configurations, the fundamentals of decoding the information on the object or phenomenon under test from a fringe pattern (AFPA – automated fringe pattern analysis), and illustrative examples of applications of interferometry in NDE. They include

K. Patorski (✉) · M. Trusiak

Department of Mechatronics, Institute of Micromechanics and Photonics,

Warsaw University of Technology, Warsaw, Poland

e-mail: [k.patorski@mchtr.pw.edu.pl](mailto:k.patorski@mchtr.pw.edu.pl); [maciek.trusiak@gmail.com](mailto:maciek.trusiak@gmail.com)

vibration testing of silicon microelements and high-accuracy in-plane displacement measurements for the analyses in experimental mechanics and material engineering.

---

## Introduction

The developments of the pulsed ruby laser by T.H. Maiman in 1960 (emitting at wavelength  $\lambda = 694.3$  nm) and CW He-Ne laser by A. Javan, W.R. Bennet, and D.R. Herriot in 1961 (emitting in the near infrared,  $\lambda = 1150$  nm) are considered as the beginning of the laser era. Research on lasers as radiation sources and the resulting applications in the last 60 years represent a spectacular example and proof of combining basic sciences with technical ones, i.e., joining the theory with practice. The revolution in science and technology introduced by laser techniques has its origin in laser radiation properties being so different from the properties of thermal light sources. From the optical metrology point of view, this statement concerns, first of all, the degree of temporal and spatial coherence of radiation and small laser beam divergence. On the other hand, technological applications benefit from very high surface and spectral power densities in a quite broad spectral range. Additionally, very short and high-power pulses can be generated.

Availability of quasimonochromatic radiation (operation in a single or in a few longitudinal modes) accompanied by high spatial coherence (operation in a single transverse mode  $TEM_{00}$ , i.e., the Gaussian beam) and controlled polarization are the main reasons of dynamic developments of laser interferometry starting from the beginning of the 1960s. Superposition of two mutually coherent disturbances with fixed phase relationship enables one to obtain precise information about the complex amplitude of one of the beams with the second one treated as the reference. Progress in interferometry is obviously aided by spectacular developments in optoelectronics (new light sources and detectors), fiber technology, etc., for the needs of industries with large R&D budgets, for example, telecommunication and multimedia. As a result, modern, automated, compact measurement systems with computer processing of acquired data are widely met in various fields of science and technology (see, e.g., Cloud (1995), Patorski and Kujawinska (1993), and Williams (1993)).

Attractivity in interferometry follows from non-contact and high-accuracy inspection of a wide range of objects (from macro- to microscale, static, and dynamic, with rough and well-polished surfaces) using automated measurement processes. In this chapter, we focus on full-field interferometry with simultaneous acquisition of data from the whole area of the object under test (in contrast to point-wise scanning-based techniques). The acquired data in the form of two-dimensional intensity distribution consist of a fringe pattern, i.e., interferogram, correlogram, holographic interferometry fringes, moirégram, or structured illumination pattern. Each image has unique features depending on the measurement method. In the majority of applications, the desired information is encoded in the pattern phase distribution, i.e., the shape, spatial period, and orientation of fringes. The amplitude

modulation of fringes (related to their contrast or visibility) might encode valuable information as well. Quantitative analysis of fringe patterns is conducted using automatic fringe pattern analysis (AFPA) methods being under continuous and dynamic development motivated by current challenges (Schwider 1990; Servin et al. 2014). Over the years, various advancements in algorithmic solutions and numerical methods have been smoothly transferred to AFPA to extend its scope and increase efficiency.

This chapter is devoted to selected applications of interferometry in NDE. First, two-beam interference systems are classified, and simple mathematical description of information they provide is given. The considerations are limited to testing specular reflection or nonscattering transmission objects. This choice has been dictated by the chapter's limited extent. Light-scattering objects are treated, for example, in the chapter on ► [Chap. 13, “Shearography”](#) of this handbook. After the fundamentals, the issue of quantitative and automated analysis of fringe patterns is shortly introduced, and a robust method for effective processing of complex patterns is presented. Examples of powerful applications of interference methods in NDE highlight and complete the chapter.

---

## Encoding Information Using Two-Beam Interference

The theoretical description will be focused on the object phase information, encoding and decoding providing such measurands as three-dimensional object shape, refractive index distribution, wave front deformation, etc. Retrieving the measurand from the fringe pattern amplitude modulation will be illustrated by one of the selected applications. The object phase information can be acquired using three main methods:

1. Interference of the object beam, reflected or transmitted by the object under test with a reference beam (mostly with perfectly plane or spherical wave fronts)
2. Interference of the object beam with its displaced replica (lateral, radial, or azimuthal shear interferometry)
3. Interference of the object beam with its conjugate, i.e., the beam with the wave front representing a mirror reflection of the object beam with respect to the plane perpendicular to the beam direction of propagation, taken here as  $z$

In the above mentioned methods, the two interfering beams can:

- (a) Propagate mutually parallel (uniform field or zero fringe detection mode)
- (b) Propagate mutually inclined (finite fringe detection mode with reference or carrier fringes)

To simplify and unify the mathematical description of the three methods, the following assumptions are made:

- The amplitudes of two beams are the same and equal to unity.
- The reference beam has a plane wave front.
- The mathematical representation of the object beam in the detection (e.g., CCD matrix) plane is  $\exp\{i\varphi(x, y)\} = \exp\{i(2\pi/\lambda)f(x, y)\}$ , where  $\varphi(x, y)$  is the phase distribution in this plane,  $f(x, y)$  denotes the optical path introduced by the object under test, and  $\lambda$  is the light wavelength.
- The interfering beams are co-phasial at the origin of the detection plane coordinate system.

## Interferometry with a Reference Beam

The simplest measurement data interpretation is encountered when using a reference beam with a plane wave front. When both beams propagate along the  $z$  axis, the intensity distribution in the detection plane can be written as

$$\begin{aligned} I(x, y) &= |1 + \exp\{i\varphi(x, y)\}|^2 = [1 + \exp\{i\varphi(x, y)\}][1 + \exp\{-i\varphi(x, y)\}] \\ &= 2[1 + \cos\varphi(x, y)]. \end{aligned} \quad (1)$$

The interferogram intensity distribution represents the contour map of investigated phase disturbance  $\varphi(x, y)$ . The separation between adjacent interference fringes (contours mapping the departure from the interference field with uniform intensity distribution) corresponds to the differences of phase and optical path equal to  $\pi$  and  $\lambda/2$ , respectively, in the case of reflecting object. For testing in transmission mode, those two values double.

Let the object beam still propagate along the  $z$  axis and the reference beam propagate inclined at a small angle  $\theta$  with respect to that axis in the  $x, z$  plane. The intensity distribution in the detection plane becomes

$$\begin{aligned} I(x, y) &= |\exp\{i(2\pi/\lambda)x\sin\theta\} + \exp\{i\varphi(x, y)\}|^2 \\ &= 2\{1 + \cos[(2\pi/\lambda)x\sin\theta - \varphi(x, y)]\} = 2\{1 + \cos[(2\pi x/(\lambda/\sin\theta)) - \varphi(x, y)]\}. \end{aligned} \quad (2)$$

In the special case of interfering beams with plane wave fronts, i.e.,  $\varphi(x, y) = 0$ , rectilinear equidistant fringes parallel to the  $y$  axis are obtained. They correspond to reference (carrier) fringes with spatial period equal to  $\lambda/\sin\theta$ . Substituting  $\varphi(x, y) = (2\pi/\lambda)f(x, y)$ , we get

$$I(x, y) = 2\{1 + \cos(2\pi/\lambda)[x\sin\theta - f(x, y)]\}. \quad (3)$$

It is seen that in the case of coding phase information with inclined reference beam, we obtain fringes with local departure from straightness proportional to the optical path introduced by the object under test.

## Interferometry with a Displaced Replica of Test Beam

To avoid the necessity of producing a reference beam with perfectly flat or spherical wave front, the method using a replica of the wave front under test was developed. Its most common form is lateral shear interferometry with lateral displacement between the two interfering beams. When two assumptions are met, i.e.,

1. The tested beam undergoes slow wave front changes,
2. The lateral displacement between the two beams is small,

this method provides fringes encoding information about the first derivative of the object beam phase. Since one of the chapters of this handbook is devoted to ► [Chap. 13, “Shearography”](#), a method widely used in experimental mechanics, we will not include this measurement technique in this chapter.

## Interferometry Using Conjugate Wave Front Beams

Two beams carrying conjugate information about the phase distribution  $\varphi(x, y)$  and propagating along the  $z$  axis generate an interferogram which can be described as

$$I(x,y) = |\exp\{i\varphi(x,y)\} + \exp\{-i\varphi(x,y)\}|^2 = 2[1 + \cos 2\varphi(x,y)]. \quad (4)$$

We see that this method provides double sensitivity of the object phase detection in comparison with the first described method, i.e., two-beam interference with reference beam; see Eqs. (1) and (2). This type of interference is employed in so-called moiré interferometry described in the latter part of this chapter. It is mainly used for material testing and experimental mechanics evaluations. Conjugate beams are provided by two lowest conjugate orders +1 and -1 of a diffraction grating fixed to the object under test.

---

## Information Decoding: Automatic Fringe Pattern Analysis

Full-field interferometric methods in NDE encode the information about the measurand in fringe pattern, i.e., characteristic intensity distribution captured by a camera. Fringes are very powerful and capable information carriers. Fringe pattern generation by two-beam interference was described in details in the previous section using optical complex amplitude and intensity domain. Throughout this section, the following straightforward mathematical model of fringe patterns will be used:

$$I(x,y) = b(x,y) + a(x,y) \cos \varphi(x,y) + n(x,y), \quad (5)$$

where  $b$  denotes the background term,  $a$  denotes amplitude modulation, and  $n$  denotes the noise part.

Generally, the background term has low spatial frequency and is structured by the laser beam spatial transverse energy distribution, i.e., Gaussian. Its minimization would concern some examples of high-pass filtering of a single fringe pattern case or subtraction of two phase-shifted ones. Noise can have various origins (coherent speckling, object surface imperfections, diffraction on dust, low-quality camera, etc.). It can be treated, however, as high-frequency intensity fluctuations.

Two information carrying components of interest are the amplitude and phase modulation distributions. Automatic fringe pattern analysis techniques tend to calculate those factors out of the intensity distribution (Malacara et al. 1998; Malacara 2007; Robinson and Reid 1993; Schwider 1990; Servin et al. 2014); it is often called amplitude and phase demodulation. This process is the core of optical measurement as it computationally decodes the measurand physically encoded in a fringe pattern by the interferometric NDE methods. Spurious components, i.e., background and noise, are to be dissected and neglected to minimize their negative influence on informational terms calculation. This special kind of fringe pattern filtering constitutes its pre-processing and is employed to enhance fringe pattern analysis outcomes, hence augmenting the accuracy of full-field interferometric NDE techniques. It is important to note that full-field incoherent non-interferometric techniques play a crucial role in NDE, e.g., moiré techniques (Patorski and Kujawinska 1993) and fringe projection (structured illumination) methods (Gorthi and Rastogi 2010).

The history of fringe pattern analysis began with visual inspection of the number of fringes estimating the amount of phase difference as fringes depict a local contour map of it. Fringes were counted, i.e., a skeletonizing procedure thinned the fringes, and prior knowledge about the object or phenomena under study enabled their numbering. This procedure started a family of methods named intensity domain fringe analysis techniques realizing fringe contouring/tracking. After successful numbering phase demodulation was performed merging numbered contours into a piecewise continuous map of the measurand. The main drawback of this type of fringe analysis is low lateral resolution and poor sampling conditions for fringe skeletonizing. On the other hand, the nature of the single-frame analysis comprises the most significant advantage which is a driving force for development both in terms of fringe-tracking techniques and phase map interpolation (Zhang et al. 2002; Zhang and Guo 2014).

Information about local direction (“polarity”) of phase change is lost in the contour map due to intensity-only recording and can be retrieved by demodulating the fringe phase distribution. In this spirit, a family of so-called phase domain techniques emerged. It started in the 1970s with phase-shifting interferometry (PSI) (Bruning et al. 1974) which bloomed richly over the years (Malacara et al. 1998; Malacara 2007; Robinson and Reid 1993; Servin et al. 2014; Schwider 1990).

In PSI a number of phase-shifted interferograms (at least three but not limited in the upper band) are recorded:

$$I_i(x,y) = b_i(x,y) + a_i(x,y) \cos [\varphi(x,y) + \delta_i] + n_i(x,y), \quad (6)$$

where  $i = 1, 2, 3, \dots$  denotes the number of the interferogram and  $\delta_i$  denotes the phase step value of interferogram number  $i$  measured with respect to the first one. When the phase step is set to  $\pi/2$  and the number of interferograms to 4, the following phase-shifting phase demodulation formula applies:

$$\varphi(x,y) = \tan^{-1}[I_4 - I_2]/[I_1 - I_3]. \quad (7)$$

Phase is obtained mathematically using the inverse tangent function; therefore, its distribution is wrapped in modulo  $2\pi$  form. Phase unwrapping is needed to generate continuous phase map adding  $k*2\pi$  where  $k$  denotes the locally varying phase-unwrapping offset (Ghiglia and Pritt 1998; Martinez-Carranza et al. 2017). After unwrapping, the continuous phase distribution is to be rescaled to finalize the measurand  $f(x, y)$  decoding:

$$f(x,y) = \lambda\varphi(x,y)/4\pi. \quad (8)$$

Generally, there are several requirements to be met when acquiring a phase-shifting sequence of interferograms (e.g., displacing reference mirror axially): the phase step amount has to be precise; background, amplitude, and phase terms should be constant in time and interferograms recorded linearly. Each requirement, when not fulfilled, constitutes an independent phase measurement error source, correctable upon especially tailoring the phase-shifting algorithm formula (Malacara et al. 1998; Malacara 2007; Robinson and Reid 1993; Servin et al. 2014; Schwider 1990; Wang and Han 2004). Phase shifting is directly applicable to any interferometric full-field NDE method with preferred null-field detection mode over the finite fringe one.

State-of-the-art multi-frame phase-shifting algorithms provide the optical path measurement extreme accuracy defined in single nanometers regardless of the object characteristic (phase function shape). However, they are not applicable in the dynamic regime testing. It is dictated by the sole necessity to capture more than one interferogram for a single “instance” of the examined object/phenomenon. This feature severely limits the time resolution of the phase-shifting approach. Attempts were made to parallelize phase-shifting sequence recording using pixelated phase masks and simultaneous phase calculation (Millerd et al. 2004) from four phase-shifted intensity pixels. However, lateral resolution of this technique is significantly reduced (by a factor of 4).

The need for real-time fringe pattern analysis sparked the advent of the family of so-called single-frame methods. Historically first and up to now, the most important and popular single-frame fringe analysis technique was proposed in 1982 by Mitsuo Takeda introducing the Fourier transform (FT) into optical metrology (Takeda et al. 1982). The real interferogram was transformed into the Fourier domain, filtered globally, and transformed back to the image domain in complex form with easy access to phase (angle of complex number) and amplitude (magnitude of complex number) using arc tangent and modulus calculations. The most important feature of the interferogram tailored for single-frame processing is high spatial carrier frequency resulting in spectral separation of the autocorrelation background term

and informational side lobes. Other integral transform-based methods use localized relatives of the FT, i.e., windowed Fourier transform (Kemao 2004), continuous wavelet transform (Wang and Ma 2006; Gdeisat et al. 2006; Pokorski and Patorski 2012, 2013), and S-transform (Zhong et al. 2013).

Interestingly PSI was converted into a single-frame technique by introducing a certain linear tilt into the phase distribution resulting in pixel-to-pixel phase shift rather than interferogram-to-interferogram one. The technique, named spatial carrier phase shifting (SCPS), can be considered as accelerated PSI and is a fairly popular method for single-frame fringe analysis (Creath and Schmit 1996). Another important group of single fringe pattern analysis techniques called regularized phase tracking emerged in 1997 (Servin et al. 1997) and blossomed further on (Kai and Kemao 2013; Legarda-Sáenz et al. 2002; Servin et al. 2001; Wielgus et al. 2014). The fringe pattern is modeled by a functional, which is minimized in a regularized fashion resulting in pixel-by-pixel estimation of the continuous phase map. No carrier frequency is needed in this case, but robustness to fringe pattern imperfections and peculiarities comes with a price of enlarged functional with costly optimization (both in terms of computational load and processing time).

There was a niche in single-frame fringe pattern analysis for a method able to provide fast and accurate single-frame phase (and amplitude) demodulation with robustness to cumbersome fringe shape (significant local period and orientation variation) and strong background/noise presence. The Hilbert-Huang transform (HHT)-based fringe analysis was recently proposed to provide new means for efficient information decoding in optical metrology. It is based on the notion of 1D HHT proposed originally by Huang et al. (1998). The idea behind this signal analysis framework is to decompose it into subsignals oscillating in scales characteristic to the analyzed signal in a data-driven manner. This sifting process is called empirical mode decomposition (EMD), and it provides efficient representation of nonstationary and nonlinear signals into a set of their zero-mean-value sub-components (empirical modes) ready for accurate Hilbert transformation to create complex analytic signals with straightforward access to instantaneous phase and amplitude of each subsignal. The first empirical mode contains locally the highest signal frequencies, whereas consecutive empirical modes store locally decreasing frequency content. Therefore, we should look for noise in the first modes and for trend in the last ones.

The EMD idea was successfully extended into two dimensions to provide capable means for image processing and analysis (Bhuiyan et al. 2009; Nunes et al. 2003). At this point, it attracted attention of the optical metrology community with the first applications of BEMD (bidimensional EMD) to digital speckle pattern interferometry fringe denoising (Bernini et al. 2008; Zhou and Li 2011), fringe pattern background removal (Zhou et al. 2012), and amplitude demodulation (Bernini et al. 2009; Wielgus and Patorski 2011). The latter is performed employing 2D Hilbert spiral transform (Larkin et al. 2001; Wielgus and Patorski 2011) on the sum of the chosen empirical modes. Fringe pattern filtering can be seen as summing up informative empirical modes:

$$\begin{aligned} \text{IF}(x,y) &= \text{BIMF}_k(x,y) + \text{BIMF}_{k+1}(x,y) + \dots + \text{BIMF}_p(x,y) \\ &= a(x,y)\cos\varphi(x,y), \end{aligned} \quad (9)$$

where IF denotes the filtered interferogram,  $k$  is the number of first informative mode (every lower mode is treated as a part of the noise factor  $n(x, y)$  and discarded), and  $p$  is the number of the last informative mode (every higher mode is treated as a part of background term  $b(x, y)$  and discarded). The Hilbert spiral transform is implemented in the Fourier domain multiplying the fringe pattern spectrum by the spiral phase function (SPF). It generates a 2D complex analytical fringe pattern (Felsberg and Sommer 2001; Larkin et al. 2001) with the real part defined as pre-processed (filtered, IF) fringe pattern and the imaginary part (quadrature interferogram, IQ) as its Hilbert spiral transform:

$$\text{IQ}(x,y) = -i\exp(-i\beta)F^{-1}\{\text{SPF}^*\text{F}[\text{IF}(x,y)]\} = -a(x,y)\sin\varphi(x,y), \quad (10)$$

where  $F^{-1}$  and  $F$  denote the inverse and forward Fourier transforms and  $\beta$  denotes the local fringe orientation map. Phase demodulation needs additional calculation of the local fringe orientation map  $\beta$ , whereas amplitude demodulation can be performed directly. Complex analytical fringe pattern  $\text{IF} + i\text{IQ}$  ensures straightforward access to the local phase and amplitude distributions of interest using simple formulae:

$$\varphi(x,y) = \text{angle}(\text{IF} + i\text{IQ}) = \tan^{-1}(\text{IQ}/\text{IF}), \quad (11)$$

$$a(x,y) = \text{abs}(\text{IF} + i\text{IQ}) = \sqrt{(\text{IF}^2 + \text{IQ}^2)}, \quad (12)$$

where ‘‘abs’’ denotes modulus operation and ‘‘sqrt’’ indicates squared root calculation.

The main drawback limiting the cavalier use of BEMD was its long computational time caused by costly interpolation of its irregular extrema grid. Decomposition was accelerated employing fast estimation rather than interpolation (Bhuiyan et al. 2008) in the FABEMD algorithm (fast and adaptive BEMD). It was introduced to fringe pattern analysis (Patorski et al. 2011; Trusiak et al. 2012) and further speeded up (Trusiak et al. 2014) and went through a number of tailored specialized modifications to aid intriguingly challenging phase demodulation in various interferometric full-field techniques (Picazo-Bueno et al. 2018; Trusiak et al. 2013, 2016a, b). Activity in amplitude demodulation is also worth mentioning (Patorski and Trusiak 2013; Patorski et al. 2014; Trusiak et al. 2018).

For fringe pattern pre-processing, alternatively to EMD, a technique called variational image decomposition was recently used (Cywińska et al. 2018; Zhu et al. 2013, 2014). Its inherent feature comprises sparse decomposition of the fringe pattern into three components: noise, background, and fringes with easy access to the filtered fringe term without the need for any modes summation. The fringe pattern pre-processing technique actively supports two-frame phase-shifting techniques – a novel branch of the PSI family (Ma et al. 2014; Saide et al. 2017; Trusiak and

Patorski 2015; Vargas et al. 2012; Wielgus et al. 2015) – combining advantages of the multi-frame (accuracy and robustness to fringe shape) and single-frame analysis (dynamic regime).

---

## Time-Average Interference Microscopy for Microelement Vibration Testing

Competitive trends to miniaturize microelectromechanical systems (MEMS), used in aerospace, telecommunication, computers, automobiles, etc., introduce unprecedented requirements on their design. To meet the challenges, new materials and structural designs are being developed. Complex structures of micromechanical assemblies and electronic packages must withstand wide ranges of mechanical, thermal, or electrical loads. To ensure their optimum design, functionality, and reliability, hybrid experimental/numerical methods of analysis are implemented (Pryputniewicz 1994). They include:

- Forming theoretical models of structure behavior and implementing them into FEM (finite element method) analysis
- Incorporating knowledge about accurate geometry and mechanical properties of materials of objects being modeled, e.g., Young's modulus, fracture strength, residual stresses, etc. (Yi and Kim 1999)

Considering experimental methods which may be unified with FEM for micro-scale elements (e.g., microactuators and microsensors), it can be recognized that the conventional procedures involving strain gauges, photoelasticity, mechanical probing, etc. are generally not applicable to these measurements. The smaller the components become, the more challenging is their performance evaluation with high accuracy. Quite often, special elements such as micromembranes, micro-bridges, and microbeams (cantilevers) are used for testing purposes, both under static and dynamic load conditions. The most popular experimental techniques include the membrane bulge test, beam bending test, and frequency response evaluation. Optical full-field measurement methods, e.g., interferometry with simultaneous acquisition of experimental data from the whole surface of the tested object, combined with parallel data processing, are well suited to evaluate microscale objects because of their non-contact and noncontaminating character.

Vibration testing provides material properties at the microlevel, validation of designs, and information for optimization of the manufacturing process. The methods for resonance frequency determination and mode shape investigation can be divided into two main groups: point scanning and field methods (Bosseboeuf and Petitgrand 2003). The first group includes deflectometry, laser Doppler vibrometry, and heterodyne and homodyne interferometry. They require point-by-point scanning. The vibration phase information is usually lost. Among the field methods, time-averaged and stroboscopic interferometry techniques can be quoted. They are based on classical interferometry (Petitgrand et al. 2001) and holographic

(conventional and digital) and speckle interferometry (Pryputniewicz and Stetson 1989; Stetson 1984) and can be used for both specular and diffuse reflection objects.

Here we would like to present the state-of-the-art method for obtaining, processing, and analyzing fringe patterns in time-averaged two-beam interferometry for active silicon microelement testing. Figure 1 shows the experimental setup schematic of the Twyman-Green interferometer with the temporal phase-shifting hardware (the object under test is placed on a PZT displacement microscope stage) for interferogram processing (Patorski et al. 2004; Salbut et al. 2003). Vibrations are excited using properly placed and mounted electrodes connected to a signal generator.

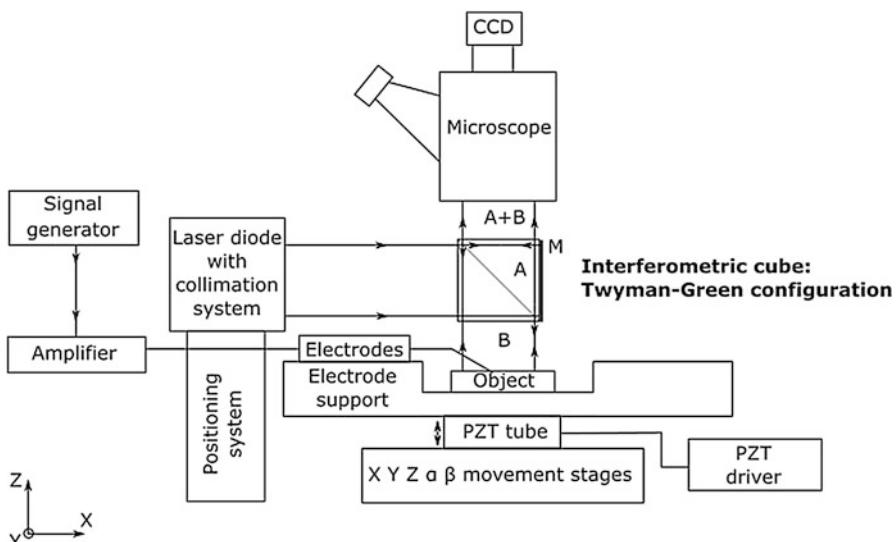
The mathematical description of the method refers to the intensity distribution of the reference beam-type two-beam interferogram, Eq. (1). It can be rewritten as (Patorski et al. 2004; Salbut et al. 2003)

$$I_{stat}(x,y) = K(x,y)\{1 + C_{stat}(x,y) \cos \varphi_{stat}(x,y)\}, \quad (13)$$

where  $K(x, y) = A^2(x, y) + B^2(x, y)$  expresses the sum of intensities of the object and reference beams (interferogram bias) – the amplitudes of the beams are denoted as  $A(x, y)$  and  $B(x, y)$ .

The contrast of interference fringes is expressed as

$$C_{stat}(x,y) = 2A(x,y)B(x,y)/[A^2(x,y) + B^2(x,y)]. \quad (14)$$



**Fig. 1** Scheme of the Twyman-Green interferometer system for silicon microelement vibration testing

The phase of nonvibrating object interferogram is given by

$$\varphi_{stat}(x,y) = (2\pi/\lambda)\text{OPD}(x,y), \quad (15)$$

where  $\text{OPD}(x,y)$  is the optical path difference between the reference and object wave fronts. It includes the influence of the mean position and non-flatness of the object surface;  $\lambda$  is the light wavelength. The index “stat” refers to stationary (nonvibrating) object.

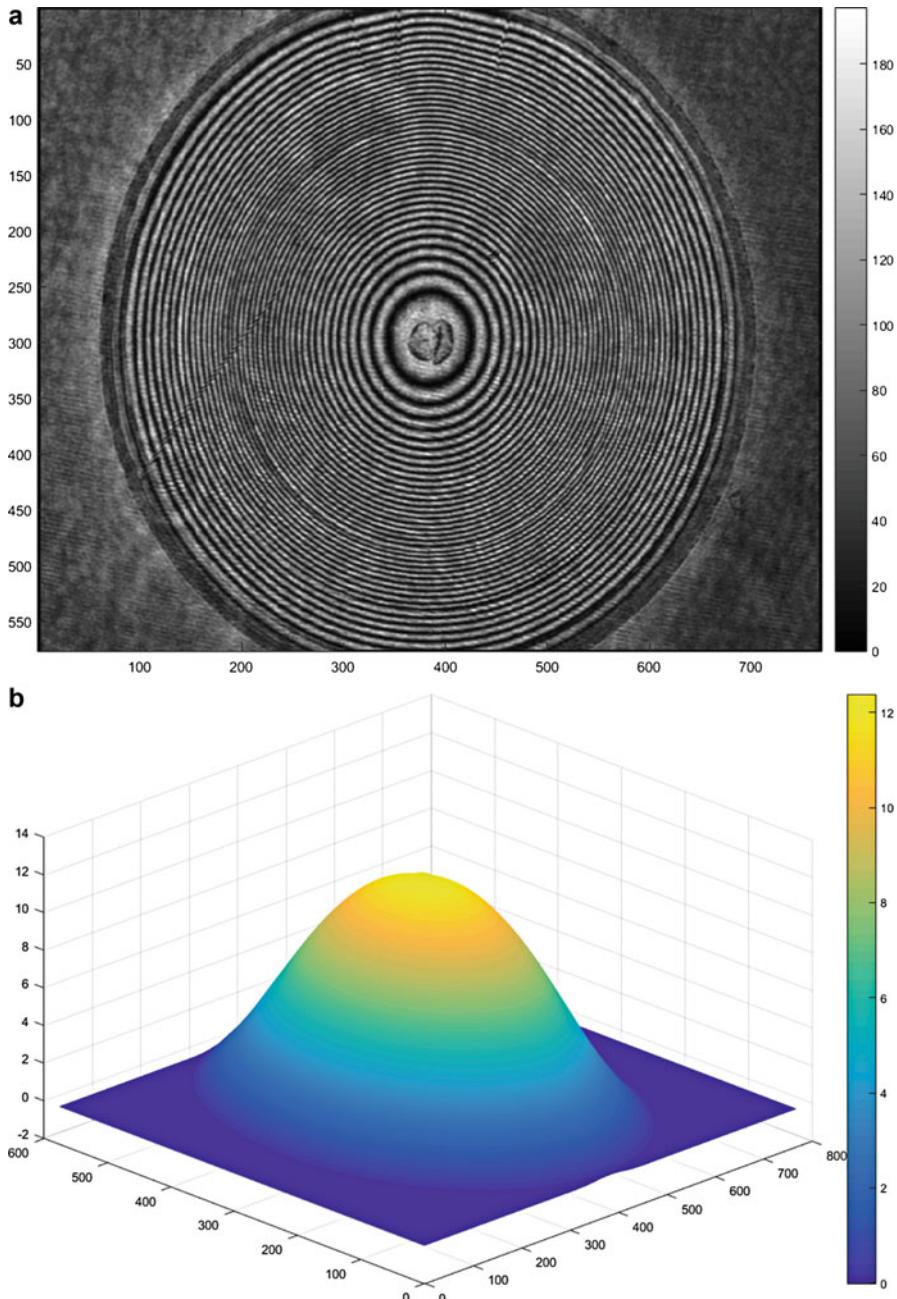
Figure 2a shows experimentally the recorded two-beam interferogram of a nonvibrating circular active silicon membrane of diameter equal to 1 mm (Salbut et al. 2003). Characteristic Newton interference fringes encode the membrane considerable departure from flatness equal to approximately 12  $\mu\text{m}$  at the membrane center. The single-frame fringe pattern processing method described in section “Information Decoding: Automatic Fringe Pattern Analysis” yields the membrane shape presented in the 3D plot, Fig. 2b.

In the case of harmonic object vibration, the interferogram intensity distribution recorded by the time-averaged method (exposure time much longer than the vibration period) can be written as (Bosseboeuf and Petitgrand 2003; Petitgrand et al. 2001; Salbut et al. 2003)

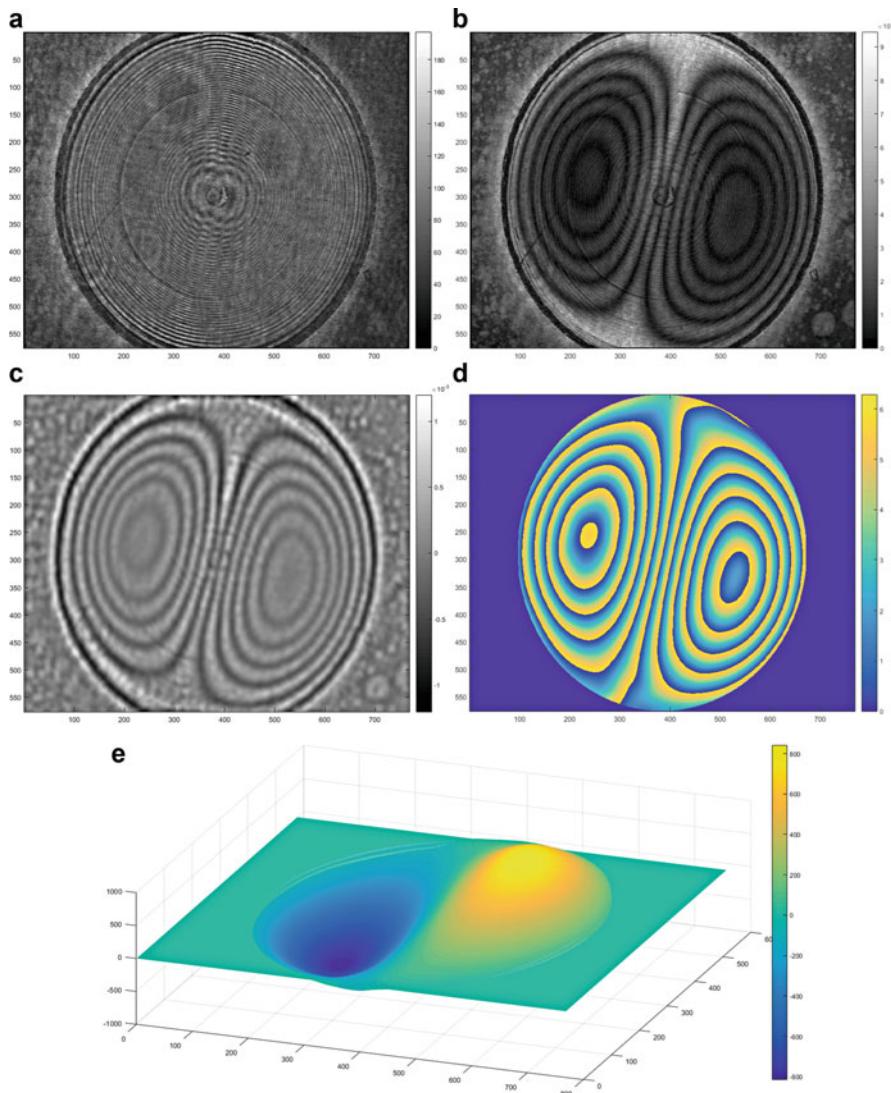
$$I_{vibr}(x,y) = K(x,y)\{1 + C_{stat}(x,y)J_0[(4\pi/\lambda)a_0(x,y)]\cos\varphi_{vibr}(x,y)\} \quad (16)$$

where  $J_0$  is the zero-order Bessel function and  $a_0(x, y)$  the vibration amplitude distribution. The sensitivity factor  $(4\pi/\lambda)$  appearing in the argument of the Bessel function relates to the experimental setup with the beam illuminating the object and observation direction along the normal to the element under test (maximum sensitivity arrangement). The phase  $\varphi_{vibr}(x, y)$  has been differentiated from  $\varphi_{stat}(x, y)$  to allow, in general, possible change of the mean position of the object under test between static and dynamic states.

Now the recorded interferogram (Newton fringes previously shown in Fig. 2a serve as carrier fringes) includes the characteristic contrast reversal bands of the carrier fringes, Fig. 3a. The amplitude (contrast) distribution of the interferogram is named as the Besselogram, Fig. 3b, and contains information on the amplitude and phase distributions of the vibration at the micromembrane resonant frequency. They can be retrieved using our advanced Besselogram processing algorithms (Patorski and Trusiak 2013; Trusiak et al. 2018). Processing requirements include, in general, short computation time and robustness to record interferogram imperfections and environmental instabilities frequently met in interference studies. The first step in Besselogram prefiltering is to obtain zero-mean-valued noise-reduced fringe pattern, Fig. 3c. Next the 2D Hilbert spiral transform (Patorski and Trusiak 2013; Trusiak et al. 2018) is applied to calculate wrapped phase fringes, Fig. 3d. After unwrapping (Martinez-Carranza 2017), the continuous vibration amplitude map is retrieved, Fig. 3e. Robustness to excited vibration mode shape, number of Bessel fringes, and its quality is noteworthy. Numerical studies and experimental evaluation clearly corroborate the advantageous features of the presented methods.



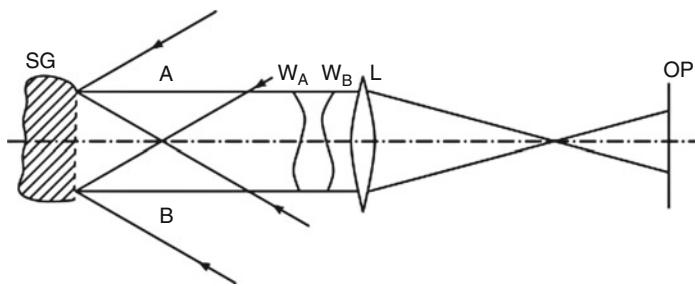
**Fig. 2** Two-beam interference fringes of a static silicon micromembrane recorded in the Twyman-Green interferometer (a) and demodulated micromembrane shape (in micrometers) using 3D representation (b)



**Fig. 3** Experimental evaluation of the time-averaged interferogram (a) of vibrating circular active silicone micromembrane: the Besselogram (b), filtered Bessel fringes (c), wrapped phase map in radians (d), and continuous full-field vibration amplitude distribution in nanometers (e)

## Moiré Interferometry for Experimental Mechanics and Material Engineering Evaluations

The so-called moiré interferometry method (Post et al. 1994; Walker 1994) represents an excellent improvement of the conventional moiré fringe method (Durelli and Parks 1970; Theocaris 1969) used for testing objects subjected to various loads.



**Fig. 4** Schematic representation of conjugate beam interferometry arrangement for in-plane displacement studies. SG, specimen grating; L, imaging optics; OP, detection plane;  $W_A$ ,  $W_B$ , wave fronts from +1 and -1 diffraction orders of SG. For simplicity one-dimensional grating case is considered

Moiré interferometry has been used for investigations of composite materials, polycrystalline materials, electronic packages in the microelectronic industry, layered materials, piezoelectric materials, fracture mechanics, structural elements and joints, biomechanics, etc. It employs high-frequency reflection type phase specimen grating SG fixed to the plane surface of the object under load, Fig. 4.

Illuminating beams A and B are mutually coherent, and when their incidence angles are tuned to the first diffraction order of the SG, the +1 diffraction order of A and -1 order of B propagate coaxially along the normal to the grating. For loaded specimen, the grating lines deviate from straightness, and the wave fronts of interfering beams are no longer plane. The complex amplitudes of the beams in the observation plane (optically conjugate to the specimen grating SG) can be described as

$$E^A_{+1}(x,y) = \exp\{i[(2\pi/d) u(x,y) + (2\pi/\lambda)w(x,y)]\}, \quad (17)$$

$$E^B_{-1}(x,y) = \exp\{-i[(2\pi/d) u(x,y) - (2\pi/\lambda)w(x,y)]\}, \quad (18)$$

where  $d$  is the spatial period of the SG whose lines are perpendicular to the plane of the drawing,  $u(x, y)$  is the in-plane displacement function corresponding to the departure of the grating lines from straightness, and  $w(x, y)$  is the out-of-plane displacement of the specimen surface. The amplitude of diffraction orders has been normalized to unity. Wave front deformations caused by in-plane displacements are equal in both diffracted beams but longitudinally reversed. On the other hand, the wave front deformations due to out-of-plane displacements are of the same value and sign in both interfering beams and therefore are eliminated by the interference. Assuming that out-of-plane displacements are such as to cause only small variations in the slope of the wave fronts, we have

$$|E^A_{+1} + E^B_{-1}|^2 = 2\{1 + \cos [(4\pi/d)u(x,y)]\}. \quad (19)$$

When the illuminating beams are not tuned to the first diffraction order of the specimen grating SG, i.e., the incidence angles of A and B are slightly smaller

or larger than that angle, carrier fringes are introduced into the interferogram. The last equation transforms into

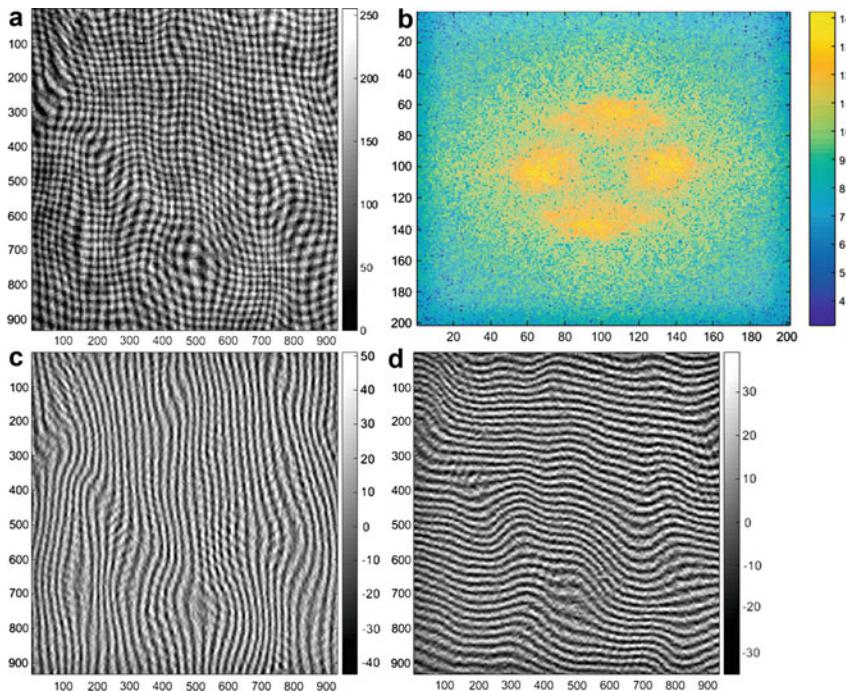
$$I(x,y) = 2\{1 + \cos [2k\theta_x x + (4\pi/d)u(x,y)]\}, \quad (20)$$

where  $|\theta_x|$  designates the angle between the diffracted conjugate beams and the grating normal (system optical axis),  $k = 2\pi/\lambda$ . The ease in introducing carrier fringes facilitates fringe processing of interferograms to expand the metrological possibilities of the method.

For the sake of simplicity, the above description of moiré interferometry has been limited to the one-dimensional case. For full-field strain analysis, we need either two mutually orthogonal in-plane displacement (or strain) fields or three separate displacement (strain) fields (Durelli and Parks 1970; Theocaris 1969). Descriptions of various optical systems implementing both approaches can be found in Post et al. (1994) and Walker (1994). Most frequently used solution is the three-mirror, four-beam moiré interferometer (Czarnek 1991).

Application of the moiré interferometry method to nondestructive material evaluation is illustrated below. Two moiré interferometry fringe patterns with encoded orthogonal in-plane displacement information  $u(x, y)$  and  $v(x, y)$  were simultaneously recorded using additive-type superposition in the stretched fabric loading experiment (Pokorski and Patorski 2012) (for the description of moiré interferometry simultaneous pattern recording, see Schmidt et al. (1997) and references therein). The resulting image is shown in Fig. 5a. In most of the image area, component fringe families are mutually quasi-perpendicular. This angular separation facilitates their extraction. However, the spectra of component fringe sets overlap partially over the image area. Furthermore, there is one area where component orthogonal fringes overlap in the way that makes visual separation difficult (see top left corner of the image). It should be emphasized that automatic fringe separation using only the Fourier transform approach would be very difficult, if not impossible, because of the mentioned spectral overlapping, Fig. 5b.

Crossed fringe pattern separation into two orthogonal fringe families can be conducted using this word should be erased a tailored HHT algorithm (Trusiak et al. 2013) or the Fourier transform method mostly rewarding the case of full spectral separation. We employed the first approach obtaining two sets of orthogonal fringes, Fig. 5c, d. They encode information about the displacement fields  $u$  and  $v$ , respectively. To calculate those maps, one can use any appropriate single-frame analysis technique described in section “[Information Decoding: Automatic Fringe Pattern Analysis](#)” with the Hilbert-Huang method chosen in this contribution. Both fringe families are phase demodulated calculating wrapped phase fringes maps, Fig. 6a, d. Then phase unwrapping is applied to generate continuous phase maps (in radians) depicted in Fig. 6b, e. We use additional 2D empirical mode filtering to smooth those phase surfaces. They have strong linear term added to facilitate single-frame analysis augmenting spatial carrier frequency which now needs to be removed. Linear function fitting in least-squared fashion and phase scaling yield nanometric displacement maps  $u$  and  $v$ , Fig. 6c, f, respectively.

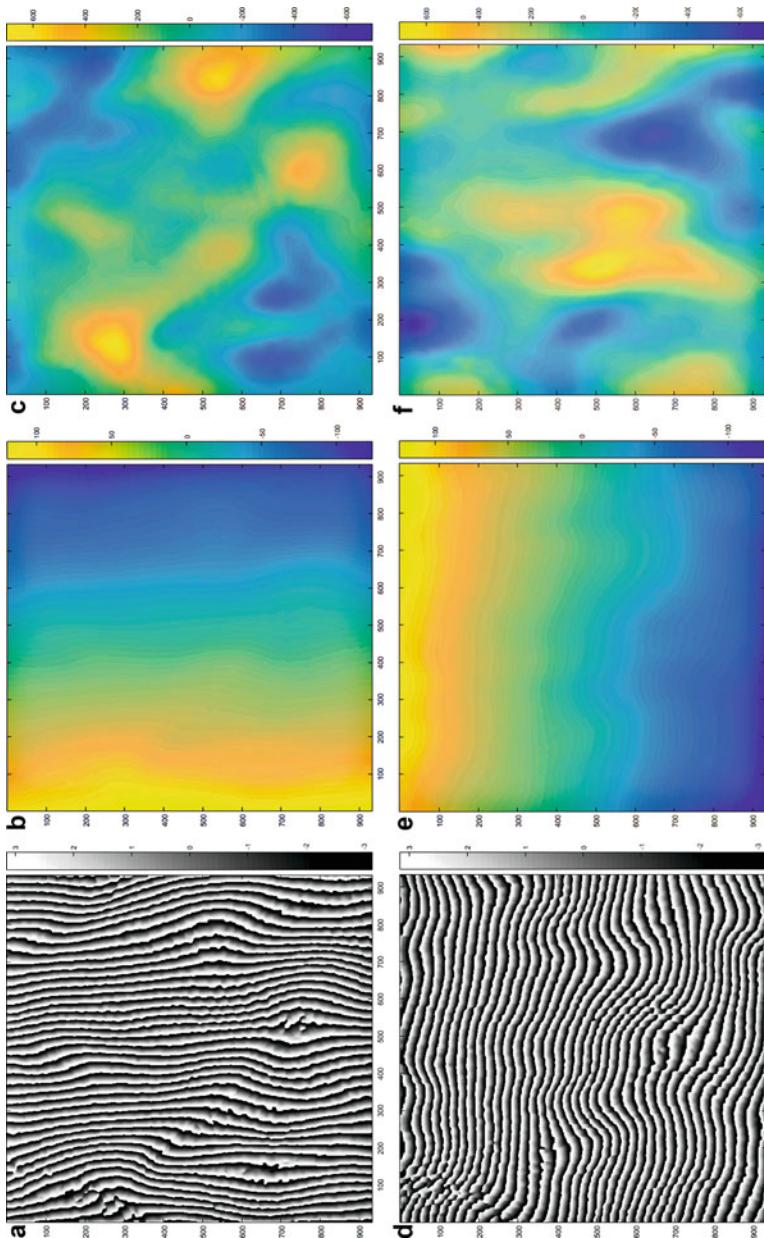


**Fig. 5** (a) Crossed interferogram obtained in the moiré interferometry experiment with simultaneously recorded, additively superimposed fringe families encoding in-plane displacement fields  $u(x, y)$  and  $v(x, y)$  in the stretched fabric experiment (Pokorski and Patorski 2012), (b) closed-up Fourier spectrum of the interferogram, (c) and (d) show both fringe sets separated using the HHT algorithm

## Summary

Optical interferometry is an indispensable tool for scientific and industrial applications in various fields. It offers unprecedented flexibility (testing and studies of static and transient events in the micro- and macroscale), real-time operation, high accuracy over adjustable measurement and sensing ranges, and considerable degree of automation. Full-field interferometry is additionally attractive providing simultaneous acquisition of data from the whole area of the tested object. By using CCD or CMOS matrix detectors, very fast parallel data processing is achieved aided by advanced interference fringe pattern processing and analysis.

For the sake of possibly comprehensive presentation of interferometric methods in NDE, within quite limited chapter extent, the following content was assumed: basic mathematical description of fundamental two-beam interference configurations; encoding and decoding the measurand or flaw information carried in the phase and/or amplitude of the interference fringe patterns; introducing the fundamentals



**Fig. 6** Results of the Hilbert-Huang transform-based algorithm phase demodulation of two fringe patterns present in Fig. 5a: wrapped phase maps (a) and (d), corresponding unwrapped phase distributions (b) and (e), unwrapped phase distributions after linear term removal and rescaling into micrometric displacement maps  $u(x, y)$  (c) and  $v(x, y)$  (f)

of quantitative, adaptive, and automated fringe pattern analysis techniques; and providing illustrative examples of applications of interferometry in NDE. The latter ones include vibration testing of silicon microelements (widely applied in variety of sensors) using time-average interference microscopy and high-accuracy in-plane displacement studies by moiré interferometry for displacement, strain, and stress analysis in experimental mechanics and material engineering. The examples presented highlight the possibilities of conducting non-contact, very high sensitivity tests of microscale objects over their quite local areas (strain concentrations) of dimensions below 1 mm. Future potential of interferometric methods in NDE is, to be true, optimistically unpredictable because of their unlimited capabilities and unpredictable development of science and technology.

**Acknowledgments** The support of National Science Center (Poland) grant OPUS 13 2017/25/B/ST7/02049 and Faculty of Mechatronics, Warsaw University of Technology statutory funds is acknowledged.

---

## References

- Bernini MB, Federico A, Kaufmann GH (2008) Noise reduction in digital speckle pattern interferometry using bidimensional empirical mode decomposition. *Appl Opt* 47(14):2592–2598
- Bernini MB, Federico A, Kaufmann GH (2009) Normalization of fringe patterns using the bidimensional empirical mode decomposition and the Hilbert transform. *Appl Opt* 48(36):6862–6869
- Bhuiyan SMA, Adhami RR, Khan JF (2008) Fast and adaptive bidimensional empirical mode decomposition using order-statistics filter based envelope estimation. *EURASIP J Adv Signal Process* 2008(164):725356
- Bhuiyan SMA, Attoh-Okine NO, Barner KE, Ayenu-Prah AY, Adhami RR (2009) Bidimensional empirical mode decomposition using various interpolation techniques. *Adv Adapt Data Anal* 01(02):309–338
- Bosseboeuf A, Petitgrand S (2003) Application of microscopic interferometry techniques in the MEMS field. *Proc SPIE* 5145:1–16
- Bruning JH, Herriott DR, Gallagher JE, Rosenfeld DP, White AD, Brangaccio DJ (1974) Digital wavefront measuring interferometer for testing optical surfaces and lenses. *Appl Opt* 13:2693–2703
- Cloud G (ed) (1995) Optical methods of engineering analysis. Cambridge University Press, Cambridge
- Creath K, Schmit J (1996) N-point spatial phase-measurement techniques for non-destructive testing. *Opt Lasers Eng* 24(5–6):365–379
- Cywińska M, Trusiaik M, Mico V, Patorski K (2018) Single-frame fringe pattern analysis using modified variational image decomposition aided by the Hilbert transform for fast full-field quantitative phase imaging. In: Proceedings of SPIE 10677, Unconventional Optical Imaging, 106772B
- Czarnek R (1991) Three-mirror, four-beam moiré interferometer and its capabilities. *Opt Lasers Eng* 13(2):93–101
- Durelli AJ, Parks VJ (eds) (1970) Moiré Analysis of Strain. Prentice Hall, Englewood Cliffs
- Felsberg M, Sommer G (2001) The monogenic signal. *IEEE Trans Signal Process* 12(49):3136–3144
- Gdeisat MA, Burton DR, Lalor MJ (2006) Spatial carrier fringe pattern demodulation by use of a two-dimensional continuous wavelet transform. *Appl Opt* 45(34):8722–8732

- Ghiglia DC, Pratt MD (eds) (1998) Two-dimensional phase unwrapping: theory, algorithms, and software. Wiley, New York
- Gorthi SS, Rastogi P (2010) Fringe projection techniques: whither we are? *Opt Lasers Eng* 48(2):133–140
- Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. *Proc R Soc Lond A* 454(1971):903–995
- Kai L, Kemao Q (2013) Improved generalized regularized phase tracker for demodulation of a single fringe pattern. *Opt Express* 21(20):24385–24397
- Kemao Q (2004) Windowed Fourier transform for fringe pattern analysis. *Appl Opt* 43(13): 2695–2702
- Larkin KG, Bone DJ, Oldfield MA (2001) Natural demodulation of two-dimensional fringe patterns. I. General background of the spiral phase quadrature transform. *J Opt Soc Am* 18(8):1862–1870
- Legarda-Sáenz R, Osten W, Jüptner W (2002) Improvement of the regularized phase tracking technique for the processing of nonnormalized fringe patterns. *Appl Opt* 41(26):5519–5526
- Ma J, Wang Z, Pan T (2014) Two-dimensional continuous wavelet transform algorithm for phase extraction of two-step arbitrarily phase-shifted interferograms. *Opt Lasers Eng* 55:205–211
- Malacara D (ed) (2007) Optical shop testing. Wiley, New York
- Malacara D, Servin M, Malacara Z (eds) (1998) Interferogram analysis for optical testing. Marcel Dekker, New York
- Martinez-Carranza J, Falaggis K, Kozacki T (2017) Fast and accurate phase-unwrapping algorithm based on the transport of intensity equation. *Appl Opt* 56(25):7079–7088
- Millerd JE, Brock NJ, Hayes JB, North-Morris MB, Novak M, Wyant JC (2004) Pixelated phase-mask dynamic interferometer. In: Proceedings of SPIE 5531, Interferometry XII: techniques and analysis
- Nunes JC, Bouaoune Y, Delechelle E, Niang O, Bunel Ph (2003) Image analysis by bidimensional empirical mode decomposition. *Image Vis Comput* 21(12):1019–1026
- Patorski K, Kujawinska M (eds) (1993) Handbook of the moiré fringe technique. Elsevier, Amsterdam
- Patorski K, Trusiak M (2013) Highly contrasted Bessel fringe minima visualization for time-averaged vibration profilometry using Hilbert transform two-frame processing. *Opt Express* 21(14):16863–16881
- Patorski K, Styk A, Sienicki Z (2004) Time-average interference microscopy for vibration testing of silicon microelements. *Proc SPIE* 6158:615806-1
- Patorski K, Pokorski K, Trusiak M (2011) Fourier domain interpretation of real and pseudo-moiré phenomena. *Opt Express* 19(27):26065–26078
- Patorski K, Trusiak M, Tkaczyk T (2014) Optically-sectioned two-shot structured illumination microscopy with Hilbert-Huang processing. *Opt Express* 22(8):9517–9527. *Virtual J Biomed Opt* 9(6)
- Petitgrand S, Yahiaoui R, Bosseboeuf A, Danaie K (2001) Quantitative time-averaged microscopic interferometry for micromechanical device vibration mode characterization. *Proc SPIE* 4400:51–60
- Picazo-Bueno JA, Trusiak M, García J, Patorski K, Micó V (2018) Hilbert–Huang single-shot spatially multiplexed interferometric microscopy. *Opt Lett* 43(5):1007–1010
- Pokorski K, Patorski K (2012) Separation of complex fringe patterns using two-dimensional continuous wavelet transform. *Appl Opt* 51(35):8433–8439
- Pokorski K, Patorski K (2013) Processing and phase analysis of fringe patterns with contrast reversals. *Opt Express* 21(19):22596–22609
- Post D, Han B, Ifju PG (eds) (1994) High sensitivity moiré: experimental analysis for mechanics and materials science and technology. Springer, New York
- Przputniewicz RJ (1994) A hybrid approach to deformation analysis. *Proc SPIE* 2342:282–296
- Przputniewicz RJ, Stetson KA (1989) Measurement of vibration patterns using electro-optical holography. *Proc SPIE* 1162:456–467

- Robinson D, Reid G (eds) (1993) *Interferogram analysis: digital fringe pattern measurement*. Institute of Physics, Bristol
- Saide D, Trusiak M, Patorski K (2017) Evaluation of adaptively enhanced two-shot fringe pattern phase and amplitude demodulation methods. *Appl Opt* 56(19):5489–5500
- Salbut L, Patorski K, Jozwik M, Kacperski J, Gorecki C, Jacobelli A, Dean T (2003) Active micro-elements testing by interferometry using time-average and quasi-stroboscopic techniques. *Proc SPIE* 5145:23–32
- Schmidt J, Patorski K, Creath K (1997) Simultaneous registration of in- and out-of-plane displacements in modified grating interferometry. *Opt Eng* 36(8):2240–2248
- Schwider J (1990) Advanced evaluation techniques in interferometry. In: Wolf E (ed) *Progress in optics*, vol 28. Elsevier, Burlington, pp 271–359
- Servin M, Marroquin JL, Cuevas FJ (1997) Demodulation of a single interferogram by use of a two-dimensional regularized phase-tracking technique. *Appl Opt* 36(19):4540–4548
- Servin M, Marroquin JL, Cuevas FJ (2001) Fringe-follower regularized phase tracker for demodulation of closed-fringe interferograms. *J Opt Soc Am A* 18(3):689–695
- Servin M, Quiroga JA, Padilla M (eds) (2014) *Fringe pattern analysis for optical metrology: theory, algorithms, and applications*. Wiley, New York
- Stetson KA (1984) Holographic vibration analysis. In: Erf RE (ed) *Holographic nondestructive testing*. Academic, New York, pp 182–220
- Takeda M, Ina H, Kobayashi S (1982) Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. *J Opt Soc Am* 72(1):156–160
- Theocaris PS (1969) Moire fringes in strain analysis. Pergamon, Oxford
- Trusiak M, Patorski K (2015) Two-shot fringe pattern phase-amplitude demodulation using Gram-Schmidt orthonormalization with Hilbert-Huang pre-filtering. *Opt Express* 23(4):4672–4690
- Trusiak M, Patorski K, Wielgus M (2012) Adaptive enhancement of optical fringe patterns by selective reconstruction using FABEMD algorithm and Hilbert spiral transform. *Opt Express* 20(21):23463–23479
- Trusiak M, Patorski K, Pokorski K (2013) Hilbert-Huang processing for single-exposure two-dimensional grating interferometry. *Opt Express* 21(23):28359–28379
- Trusiak M, Wielgus M, Patorski K (2014) Advanced processing of optical fringe patterns by automated selective reconstruction and enhanced fast empirical mode decomposition. *Opt Lasers Eng* 52(1):230–240
- Trusiak M, Mico V, Garcia J, Patorski K (2016a) Quantitative phase imaging by single-shot Hilbert–Huang phase microscopy. *Opt Lett* 41(18):4344–4347
- Trusiak M, Śluzewski L, Patorski K (2016b) Single shot fringe pattern phase demodulation using Hilbert-Huang transform aided by the principal component analysis. *Opt Express* 24(4): 4221–4238
- Trusiak M, Styk A, Patorski K (2018) Hilbert–Huang transform based advanced Bessel fringe generation and demodulation for full-field vibration studies of specular reflection micro-objects. *Opt Lasers Eng* 110:100–112
- Vargas J, Quiroga JA, Sorzano COS, Estrada JC, Carazo JM (2012) Two-step demodulation based on the Gram–Schmidt orthonormalization method. *Opt Lett* 37(3):443–445
- Walker CA (1994) A historical review of moiré interferometry. *Exp Mech* 34(2):281–299
- Wang Z, Han B (2004) Advanced iterative algorithm for phase extraction of randomly phase-shifted interferograms. *Opt Lett* 29(14):1671–1673
- Wang Z, Ma H (2006) Advanced continuous wavelet transform algorithm for digital interferogram analysis and processing. *Opt Eng* 45(4):045601
- Wielgus M, Patorski K (2011) Evaluation of amplitude encoded fringe patterns using the bidimensional empirical mode decomposition and the 2D Hilbert transform generalizations. *Appl Opt* 50(28):5513–5523
- Wielgus M, Patorski K, Etchepareborda P, Federico A (2014) Continuous phase estimation from noisy fringe patterns based on the implicit smoothing splines. *Opt Express* 22:10775–10791

- Wielgus M, Sunderland Z, Patorski K (2015) Two-frame tilt-shift error estimation and phase demodulation algorithm. *Opt Lett* 40(15):3460–3463
- Williams DC (ed) (1993) Optical methods in engineering metrology. Chapman & Hall, London
- Yi Y, Kim Ch-J (1999) Measurement of mechanical properties for MEMS materials. *Meas Sci Technol* 10(8):706–716
- Zhang Z, Guo H (2014) Principal-vector-directed fringe-tracking technique. *Appl Opt* 53(31): 7381–7393
- Zhang D, Ma M, Arola DD (2002) Fringe skeletonizing using an improved derivative sign binary method. *Opt Lasers Eng* 37(1):51–62
- Zhong M, Chen W, Wang T, Su X (2013) Application of two-dimensional S-transform in fringe pattern analysis. *Opt Lasers Eng* 51(10):1138–1142
- Zhou Y, Li H (2011) Adaptive noise reduction method for DSPI fringes based on bi-dimensional ensemble empirical mode decomposition. *Opt Express* 19(19):18207–18215
- Zhou X, Podoleanu AG, Yang Z, Yang T, Zao H (2012) Morphological operation-based bi-dimensional empirical mode decomposition for automatic background removal of fringe patterns. *Opt Express* 20(22):24247–24262
- Zhu X, Chen Z, Tang C (2013) Variational image decomposition for automatic background and noise removal of fringe patterns. *Opt Lett* 38(3):275–277
- Zhu X, Tang C, Li B, Sun C, Wang L (2014) Phase retrieval from single frame projection fringe pattern with variational image decomposition. *Opt Lasers Eng* 59(8):25–33



# Shearography

13

Lianxiang Yang and Junrui Li

## Contents

Introduction .....	384
Principles of Digital Shearography .....	386
Schematic Setups of Shearography .....	386
Formation of Fringes in Digital Shearography .....	388
Relationship Between Relative Phase Change $\Delta$ and Deformation Derivatives .....	390
Fringe Interpretation .....	393
Evaluation of Shearogram by Phase Shift Technique .....	393
Temporal Phase Shift Shearography .....	393
Spatial Phase Shift Shearography .....	398
Practical Applications of Digital Shearography for NDT .....	404
Introduction .....	404
Instrumentation .....	405
Measuring Sensitivity .....	406
Methods of Stressing .....	411
Conclusions .....	414
Potentials .....	414
Limitations .....	415
Summary .....	416
References .....	416

### Abstract

This chapter provides a review of the developments of *shearography* and its applications in nondestructive testing (NDT) and evaluation (NDE). Shearography, or speckle pattern shearing interferometry, is an interferometric technique for full-field, non-contact measurement of the first derivative of surface deformation, which is strain information. It was originally developed to overcome

---

L. Yang (✉) · J. Li

Optical Laboratory, Department of Mechanical Engineering, Oakland University, Rochester,  
MI, USA

e-mail: [yang2@oakland.edu](mailto:yang2@oakland.edu); [jli23456@oakland.edu](mailto:jli23456@oakland.edu)

several limitations of holography by eliminating the reference beam. Consequently, ***shearography*** is an interferometric technique that has very high measurement sensitivity but, through its direct measurement of strain information, is less sensitive to environmental disturbances. Therefore, it is a practical tool which can be used in field/factory settings. Furthermore, the self-reference system has a simple optical layout and balanced optical paths, which enables the construction of a very compact and practical ***shearographic*** sensor using a cost-economical diode-laser. In NDT, ***shearography*** reveals defects in an object by identifying defect-induced deformation anomalies through the display of strain concentrations (i.e., first derivatives of surface deformation). Shearography has already received considerable industry acceptance, in particular for nondestructive testing of such materials as composites and honeycomb structures. Another application of shearography is for strain measurement. This chapter focuses on the digital version of shearography for NDT. After discussion of the fundamentals of shearography, the recent developments and applications of shearography, as well as its potential and limitations, will be demonstrated through examples of NDT for different applications.

---

## Introduction

Nondestructive testing (NDT) is a group of techniques used to evaluate the properties of a structure or product without causing damage (Cartz 1995). It is a highly valuable technique in product evaluation, troubleshooting, and research since it does not alter the object being inspected. Recently, a requirement has arisen for a better NDT technique, due to a demand for greater product performance and reliability, especially for online inspection. To meet this requirement, the new NDT technique should contain such advantages as real-time inspection, whole-field test, non-contact measurement, and high-sensitivity. Most optical NDT techniques, such as holographic interferometry (Vest 1979), electronic speckle pattern interferometry (ESPI) (Løkberg 1987), shearography (Steinchen and Yang 2003), thermography (Clark et al. 2003), profilometry (Targowski et al. 2004), digital image correlation (DIC) (Li et al. 2017a), etc., have the virtues mentioned above. Of these techniques, shearography has already been applied in the field of NDT for decades and been proven to be a practical tool for online inspection due to its simple setup, direct measurement of strain information and relatively insensitive to environment interruptions. It is gaining more and more acceptance and applications in the automotive and aerospace industries for NDT/NDT, strain measurement, and vibration analysis.

Shearography, sometimes called speckle pattern shearing interferometry, is a laser-based optical interferometric method that is similar to ESPI. Unlike ESPI, which measures deformation, shearography directly measures the gradient of the deformation, e.g., strain information, using a special shearing device. Since object defects induce strain concentrations during loading, shearography is much more sensitive for detection of defects with strain anomalies than for deformation

anomalies. Also, shearography is insensitive to environmental disturbances, such as environmental vibration and motion, because the rigid-body movement does not induce strain. Moreover, shearography uses the self-reference technique, which does not require a reference beam and leads to a much more compact and stable optical setup. The self-reference also reduces the coherent length requirement of the laser and enables to use simple and cheap diode lasers for illumination.

The development of shearography can be divided into three stages: photographic shearography, electronic shearography, and digital shearography. In the early decades of the development of shearography, image recording was restricted to dry plate photography, and photographic emulsion was used as the recording media to obtain very high-quality images (Hung 1982; Sirohi 1984). However, traditional photographic shearography requires a slow and complicated wet photographic process, limiting its industrial applications. To address this limitation, a reusable thermoplastic plate was adopted to replace the traditional dry plate, enabling the image to be instantly obtained after recording (Hung and Hovanesian 1982). An initial attempt to achieve real-time inspection using shearography, the thermoplastic-based photographic shearography partially solves the existing problem of slow processing, but the inconsistent quality and the high price of the thermoplastic plate deterred its further application. The slow processing problem for real-time shearographic inspection was not fully solved until the electronic technique was introduced to shearography, known as electronic shearography (Steinchen et al. 1994). The electronic sensor allowed the interferogram to be captured electronically and then displayed on a TV monitor, enabling real-time display of the shearographic measurement result. However, the spatial resolution was much lower compared to photographic shearography due to the low resolution of the video camera.

Digital shearography as the latest version of shearography and the successor to electronic shearography, uncovered the full capability of shearography in many aspects (Hung 1999). Digital shearography utilizes a digital camera to capture information and create the digital photograph. There are two primary types of image sensors in a digital camera – complementary metal-oxide semiconductor (CMOS) and charge-coupled device (CCD) – and each has its advantages. The utilization of the digital camera provides good image resolution and quality to shearography, as well as the capability to display the result instantly. More importantly, the digitized image can be processed using multiple digital image processing algorithms as well as the phase shifting technique, which increases the measurement sensitivity of shearography by tens of times. Today, digital shearography with phase shifting technique has become the most favorable shearography version due to its ease of use, fast evaluation, high sensitivity, and versatility. Based on the phase shifting technique, digital shearography can be cataloged as either temporal phase shift digital shearography (TPS-DS) or spatial phase shift digital shearography (SPS-DS) TPS-DS is usually suitable for static or quasi-static tests, whereas the SPS-DS is preferable for dynamic tests.. Although the application of TPS-DS is limited to quasi-static applications, it provides a higher quality phase map compared to the SPS-DS. In recent decades, both the dynamic range of the TPS-DS and the phase

map quality of SPS-DS has been greatly improved with the rapid development of the electronic technique.

At the same time, with the development of phase shift digital shearography, the application of digital shearography has been greatly extended, and the setup has been optimized for various new applications. All these efforts have greatly enhanced the diversity of the NDT applications of shearography. For instance, the rubber industry evaluates the quality of tires using shearography (Hung 1996; Krivtsov et al. 2002; Baldwin and Bauer 2008; Zhang et al. 2013), and the aerospace industry utilizes shearography to examine the structure of aircraft (Yang and Hung 2004; Ibrahim et al. 2004; Lobanov et al. 2009; Růžek and Běhal 2009). In addition to these two areas, many types of research related to the NDT of composites using shearography have also been investigated and published in the same period (Toh et al. 1990; Hung 1996; Burleigh 2002; Gryzgoridis and Findeis 2008; De Angelis et al. 2012). As the NDT of composites using shearography reaches its mature stage, ASTM has published a standard for NDT of polymer matrix composites and sandwich core materials using shearography (ASTM E2581-14 2014). Furthermore, other applications of shearography includes: strain measurement (Hung 1982; Hung and Wang 1996; Steinchen et al. 1998a; Kästle et al. 1999), residual stress (Hung and Hovanesian 1990; Hathaway et al. 1997; Hung et al. 1997; Diaz et al. 2000), material properties (Lee et al. 2004, 2008; Huang et al. 2009; Taillade et al. 2011), 3D shape (Shang et al. 2000; Groves et al. 2004), vibration (Sim et al. 1995; Toh et al. 1995; Steinchen et al. 1996; Yang et al. 1998), and leak detection (Hung and Shi 1998).

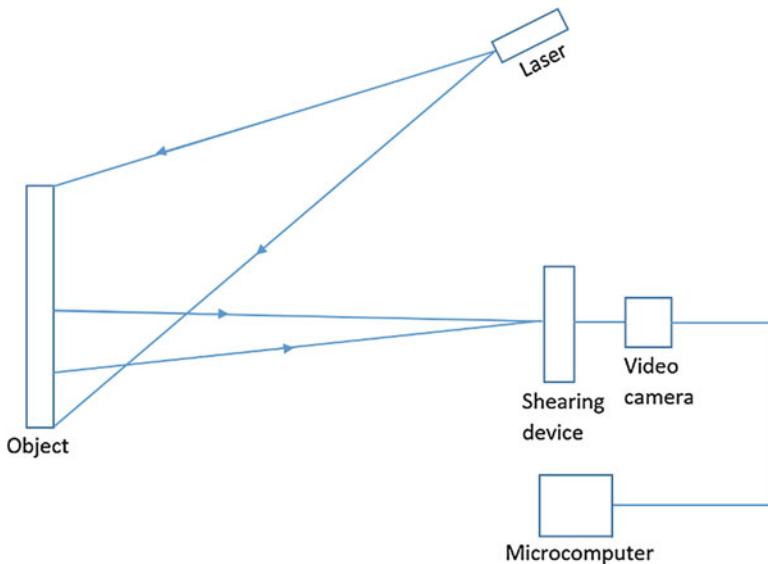
---

## Principles of Digital Shearography

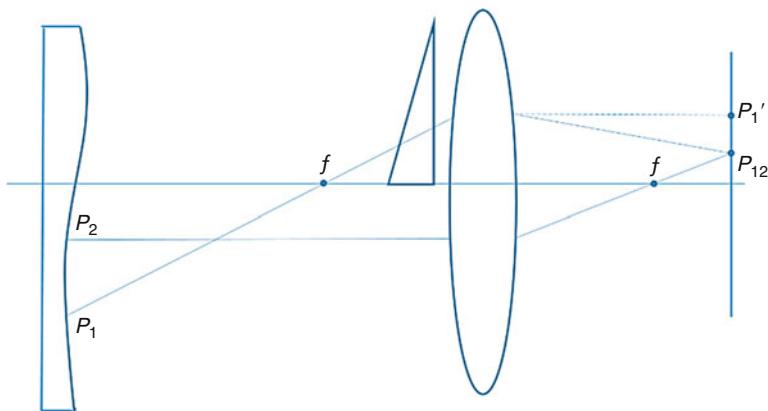
### Schematic Setups of Shearography

Several basic setups of shearography are illustrated in this section to introduce the fundamentals of shearography as well as some basic concepts, such as the shearing device, shearing amount, intensity equation, and fringe generation. The schematic of shearography is shown in Fig. 1. The test object is illuminated by an expanded laser point, and the object image is captured by the digital camera connected to a computer for fringe evaluation. A required special shearing device is placed in front of the camera to introduce image shearing. The shearing device brings two non-parallel light beams that reflected from two different object points to interfere with each other on the camera sensor. The usage of the shearing device allows shearography to be a self-referenced interferometric system, which provides a distinguishing feature compared to the other interferometric systems. Three different practical shearing devices will be described, corresponding to three well-known shearography setups.

The schematic arrangement of a very early shearography setup is shown in Fig. 2. An optical wedge is used as the shearing device in this shearographic setup (Hung 1982). The glass wedge is a small-angle ( $\sim 0.75^\circ$ ) prism that deflects rays passing

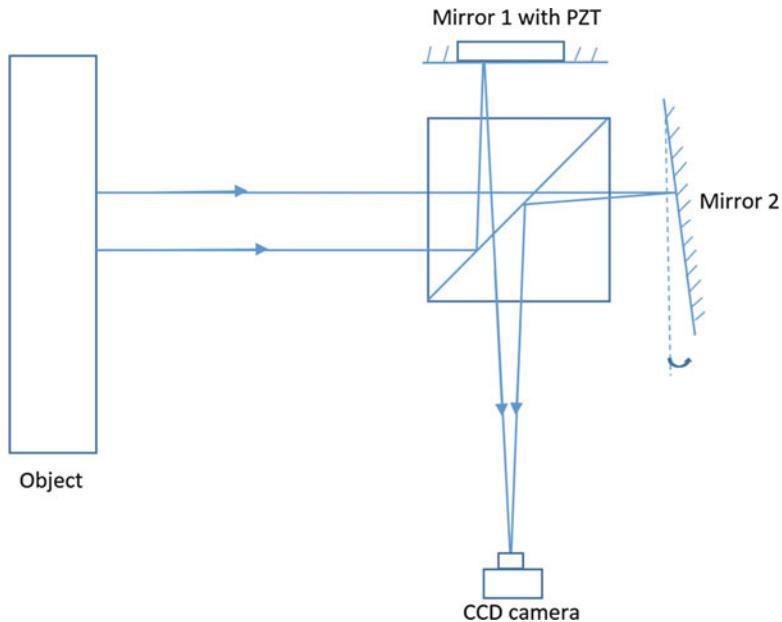


**Fig. 1** Schematic of the shearographic system



**Fig. 2** Schematic of the shearographic setup with an optical wedge

through it. The wedge is placed in front of the lens of a camera, covering one-half of the lens, and induces the image from two different points on the object surface to coincide together. The magnitude of the shearing, known as the shearing amount, is determined by the wedge angle of the shearing device, while the orientation of the wedge determines the direction of the shearing, known as the shearing direction. The shearing effect introduced by the wedge can be explained using the same figure. The shearing direction is in the  $x$ -direction, and the shearing amount is  $\delta x$  on the object surface (distance between  $P_1$  and  $P_2$ ). The rays from two different points ( $P_1$  and  $P_2$ ) meet at the same location ( $P'_{12}$ ) on the image plane.



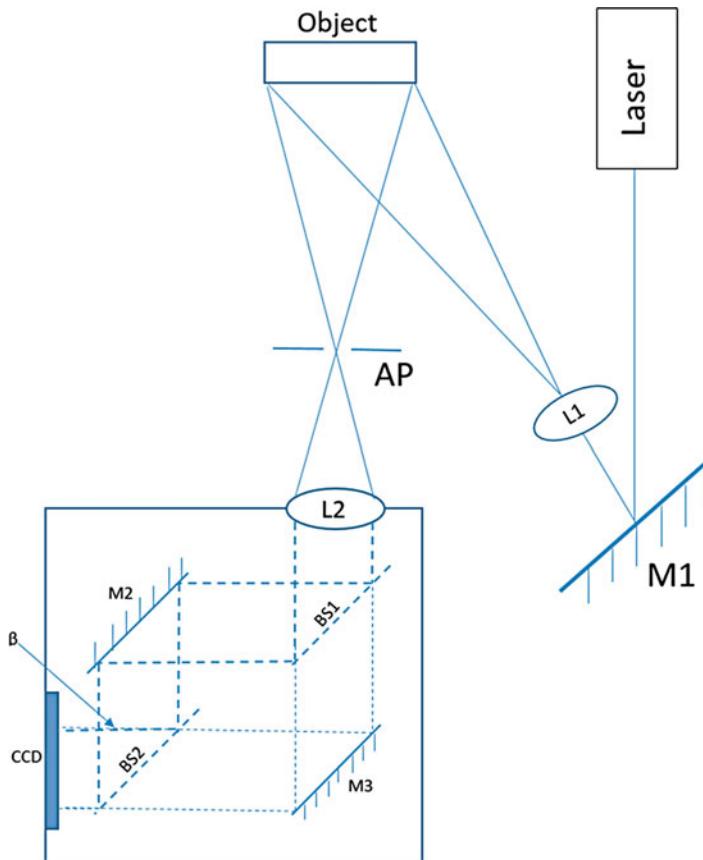
**Fig. 3** Schematic of the shearographic setup with the modified Michelson interferometer

Figure 3 shows another popular shearographic setup, which uses a modified Michelson interferometer as the shearing device (Hung 1974; Leendertz and Butters 1973). Shearing is created by tilting mirror 2 of the Michelson interferometer a very small angle. The tilting of mirror 2 brings laser rays from two different points on the object surface to one point on the sensor plane. These laser rays interfere with each other and produce the speckle pattern in the interferogram. This setup has the advantages of a simple structure and adjustable shearing direction and shearing amount. Due to these merits, this setup is widely used in the TPS-DS system and can also be used in the SPS-DS system with some limitations.

The third schematic shearographic setup, shown in Fig. 4, is based on a Mach-Zehnder interferometer. It is the first frequency carrier SPS-DS system, proposed by G. Pedrini in 1996 (Pedrini et al. 1996). The carrier frequency and the shearing amount is introduced by tilting mirror 2 in the Mach-Zehnder interferometer. Additionally, a slit aperture (AP) is placed in front of the imaging lens to control the spectral bandwidth in the spatial frequency domain. The sheared speckle pattern created by the shearing device is recorded and processed by the digital camera.

### Formation of Fringes in Digital Shearography

No matter what kind of shearing devices is used, the fundamental of shearography is to bring light waves reflected from two points ( $P_1$  and  $P_2$ ) from the object surface to meet at the same point  $P'_{12}$  in the image plane due to the



**Fig. 4** Schematic of the shearographic setup with the Mach-Zehnder interferometer

shearing device as shown in Fig. 2, where they interfere each other. Following the explanation in the previous section, the light waves reflected from two different points can be described using the exponential functions:

$$\begin{aligned} U_1 &= a_1 e^{i\theta_1} \\ U_2 &= a_2 e^{i\theta_2} \end{aligned} \quad (1)$$

Where  $\theta_1$  and  $\theta_2$  represent the random phase angle of the two points, and  $a_1$  and  $a_2$  are the amplitude of the two lights. The total light field  $U_t$  at the point  $P'_{12}$  on the image plane can be expressed:

$$U_t = U_1 + U_2 = a_1 e^{i\theta_1} + a_2 e^{i\theta_2} \quad (2)$$

The intensity of the captured image, i.e., the speckle pattern, at the point  $P'_{12}$  can be obtained by the following equation:

$$\begin{aligned}
I = U_t U_t^* &= (a_1 e^{i\theta_1} + a_2 e^{i\theta_2})(a_1 e^{-i\theta_1} + a_2 e^{-i\theta_2}) \\
&= (a_1^2 + a_2^2) + a_1 a_2 [e^{i(\theta_1 - \theta_2)} + e^{-i(\theta_1 - \theta_2)}] \\
&= (a_1^2 + a_2^2) + 2a_1 a_2 \cos(\theta_1 - \theta_2) \\
&= I_0(1 + \gamma \cos \phi)
\end{aligned} \tag{3}$$

Where  $U_t^*$  is the complex conjugate of  $U_t$ ;  $I_0 = a_1^2 + a_2^2$ , is the background;  $\gamma = 2a_1 a_2 / (a_1^2 + a_2^2)$ , is the contrast; and  $\phi = \theta_1 - \theta_2$ , is the phase difference between the rays from the two points. Figure 5a shows the intensity image of the speckle pattern before loading.

After the object is loaded, the phase difference between the two points can be expressed as  $\phi' = \phi + \Delta$ , where  $\Delta$  is the relative phase change before and after loading between the two waves from  $P_1$  to  $P_2$  and it results from the relative deformation between the two points. The intensity of the speckle pattern after loading, as shown in Fig. 5b, then can be expressed:

$$\begin{aligned}
I' &= I_0[1 + \gamma \cos(\phi')] \\
&= I_0[1 + \gamma \cos(\phi + \Delta)]
\end{aligned} \tag{4}$$

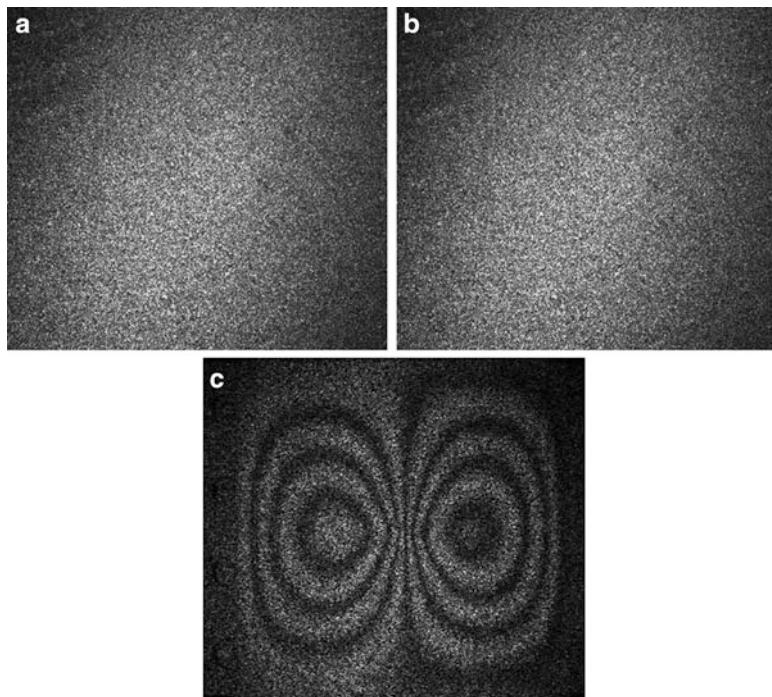
The subtraction between loaded and unloaded intensity images generated a fringe pattern, called shearogram. Because intensity cannot be negative, the absolute value of intensity change is displayed:

$$|I_s| = |I' - I| = I_0 \gamma |\cos(\phi + \Delta) - \cos(\phi)| \tag{5}$$

The dark fringe, which results when  $|I_s| = 0$ , occurs at the location where  $\Delta = 2n\pi$ , where  $n$  is the fringe order. A visible bright-dark fringe can be observed with this simple process. It is the fastest and most concise method to format the fringe without any further processing. It achieves real-time measurement and displays on a monitor controlled by a computer. Figure 5c shows a typical shearographic fringe pattern (x-direction shearing), generated using the method mentioned above, for a center-loaded square plate. The shearographic fringe pattern is butterfly-shaped and commonly called a butterfly pattern. Since the relative phase difference between two adjacent fringes is  $\Delta = 2\pi$ , the phase at each fringe can be determined by counting the fringe order.

## Relationship Between Relative Phase Change $\Delta$ and Deformation Derivatives

As explained in the previous section, the relative phase change at each fringe can be obtained by counting the fringe order. Once the relationship between the relative phase change and the relative surface deformation between the two interferometric points is established, the derivative of the deformation can be determined based on the measured relative phase change.

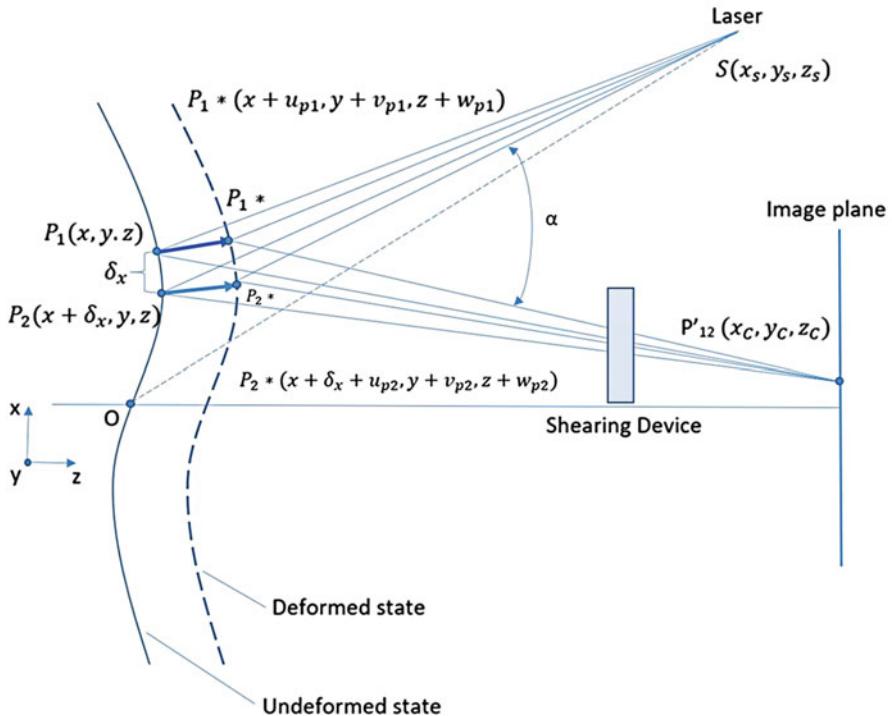


**Fig. 5** Intensity images of speckle pattern (a) before loading, (b) after loading, and (c) Shearographic fringe pattern

Assuming the two points  $P_1(x, y, z)$  and  $P_2(x + \delta x, y, z)$  on the object surface meets at the same point  $P'_{12}(x_c, y_c, z_c)$  on the sensor plane due to the shearing device, the schematic of the system is shown in Fig. 6. The relative phase change  $\Delta$  before and after loading of  $P_1$  and  $P_2$  can be expressed using the typical equation in holography (Vest 1976; Steinchen and Yang 2003), assuming the observation direction is perpendicular to the object surface, and the illumination angle lying on the XOZ plane is  $\alpha$ :

$$\begin{aligned}
 \Delta_x &= \Delta_{P1} - \Delta_{P2} \\
 &= \frac{2\pi}{\lambda} \{ [u_{P1} \sin \alpha + w_{P1}(1 + \cos \alpha)] - [u_{P2} \sin \alpha + w_{P2}(1 + \cos \alpha)] \} \\
 &= \frac{2\pi}{\lambda} [(u_{P1} - u_{P2}) \sin \alpha + (w_{P1} - w_{P2})(1 + \cos \alpha)] \\
 &= \frac{2\pi}{\lambda} [\delta u \sin \alpha + \delta w(1 + \cos \alpha)]
 \end{aligned} \tag{6}$$

where  $u_{P1}$ ,  $w_{P1}$ ,  $u_{P2}$ , and  $w_{P2}$  are the displacement of  $P_1$  and  $P_2$  in the  $x$ -direction and  $z$ -direction, respectively. If the shearing direction is in the  $x$ -direction and the distance between the two points is relatively small, the equation can be represented as:



**Fig. 6** Schematic of the interferometric effect of the shearography

$$\Delta_x = \frac{2\pi}{\lambda} \left[ \frac{\partial u}{\partial x} \sin \alpha + (1 + \cos \alpha) \frac{\partial w}{\partial x} \right] \delta x \quad (7)$$

Keeping the same illumination plane, and changing shearing direction in the  $y$ -direction, the equation becomes:

$$\Delta_y = \frac{2\pi}{\lambda} \left[ \frac{\partial u}{\partial y} \sin \alpha + (1 + \cos \alpha) \frac{\partial w}{\partial y} \right] \delta y \quad (8)$$

Similarly, when the illumination plane is in YOZ plane, Eqs. 7 (x-shearing direction) and 8 (y-shearing direction) become:

$$\Delta_x = \frac{2\pi}{\lambda} \left[ \frac{\partial v}{\partial x} \sin \alpha + (1 + \cos \alpha) \frac{\partial w}{\partial x} \right] \delta x \quad (9)$$

$$\Delta_y = \frac{2\pi}{\lambda} \left[ \frac{\partial v}{\partial y} \sin \alpha + (1 + \cos \alpha) \frac{\partial w}{\partial y} \right] \delta y \quad (10)$$

Equations 7, 8, 9, and 10 are the fundamental equations of shearography. Different shearographic setups use a different method to evaluate the relative phase change, but all of them use these four equations to calculate the derivative of deformation. For NDT, a very small or close to zero illumination angle is usually selected; under this condition, Eqs. 7, 8, 9, and 10 are simplified into two equations:

$$\Delta_x = \frac{4\pi}{\lambda} \cdot \frac{\partial w}{\partial x} \cdot \delta x \quad (11)$$

$$\Delta_y = \frac{4\pi}{\lambda} \cdot \frac{\partial w}{\partial y} \cdot \delta y \quad (12)$$

Equations 11 and 12 are usually used to interpret shearographic test results in NDT.

## Fringe Interpretation

Based on the derivation of the previous section, the measured relative phase change using shearography is directly related to the deformation gradient. In this section, the result measured by shearography will be interpreted in another aspect. Since the interpretation of holography for NDT is well-known, a comparison between holographic and shearographic measurement is presented. Figure 7 shows the results from the same NDT test measured using both holography and shearography. For this example, assume a void or delamination exists in the composite plate as the defect, and the plate is loaded by a thermal load or vacuum. The plate surface would exhibit slight buckling due to the inconsistency of material. Holography detects the flaw by detecting the deformation anomalies, and the anomaly area has a circular fringe pattern. In contrast, shearography discovers the flaw by looking for the anomalies of the first derivative of the deformation, and the defected area shows a butterfly pattern.

Figure 8 shows the different NDT results, using the phase shift technique introduced in the next section, between holography and shearography to demonstrate the concept described above. A composite plate with three de-laminated defects is inspected using the two techniques simultaneously. The holographic measurement result shows the defect with several circular anomalous fringes, and the Shearographic measurement result displays several butterfly patterns at the de-bond locations.

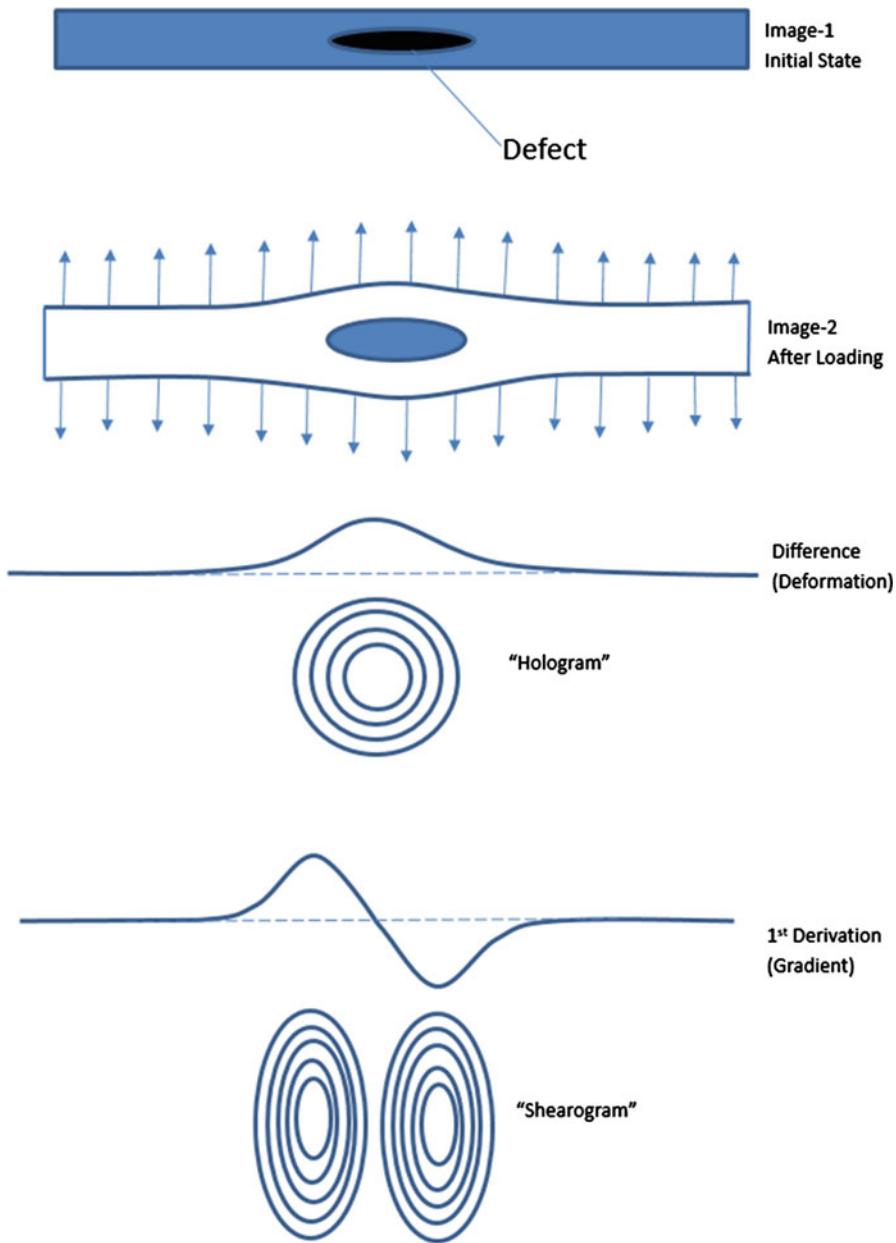
---

## Evaluation of Shearogram by Phase Shift Technique

### Temporal Phase Shift Shearography

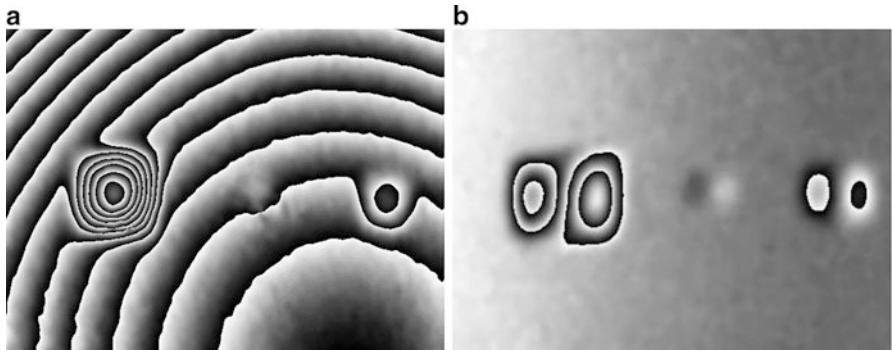
#### 4 + 4 Phase Shift Algorithm

The real-time subtraction version of digital shearography is reviewed in section “Formation of Fringes in Digital Shearography.” The fringe pattern is generated by the numerical subtraction of the interferograms captured



**Fig. 7** Schematic of the NDT result between Holography and Shearography

before and after loading. The relative phase change is calculated based on counting the fringes. However, the sensitivity of this method is limited to one fringe order, meaning that the maximum phase measurement sensitivity of real-time subtraction shearography is only  $2\pi$ . Additionally, the information



**Fig. 8** (a) NDT result obtained by Holography. (b) NDT result obtained by Shearography

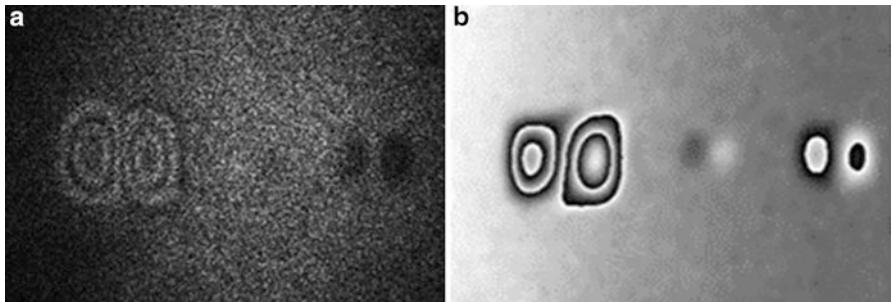
between the fringes cannot be obtained, which limits the application of real-time subtraction shearography.

To increase the phase measurement sensitivity, the phase shift technique was introduced to digital shearography. A higher phase measurement sensitivity can be achieved with the phase shift technique based digital shearography, commonly called phase shift digital shearography. The phase is directly measured in this technique rather than by counting the fringes. Phase shift digital shearography can be divided into two categories: temporal phase shift digital shearography (TPS-DS), which has the higher interferogram quality, and spatial phase shift digital shearography (SPS-DS), which has the higher dynamic range. A comparison between shearography with and without phase shift technique is shown in Fig. 9.

A typical TPS-DS setup based on the modified Michelson interferometer was shown in Fig. 3. To solve for the phase ( $\phi$  or  $\phi'$ ) using Eq. 3 or Eq. 4, at least three equations are required because there are three unknowns:  $I_0$ ,  $\gamma$ , and  $\phi$  (or  $\phi'$ ). The temporal phase shift is to record 3–4 images in a time series, and a phase shift is introduced between two images which is achieved by a piezoelectric crystal (PZT) driven mirror (Mirror 1), and the driven mirror can be moved with precise displacement to generate a known phase shift.

Though 3 equations can already solve the phase  $\phi$ , 4 equations phase shift algorithm is the most widely used in the temporal phase shift method (Steinchen and Yang 2003; Yang et al. 1995). This algorithm is fast and simple with a high phase extraction accuracy. Usually, it is called a  $4 + 4$  algorithm, i.e., 4 equations to solve  $\phi$  (phase before loading) and the other 4 to solve  $\phi'$  (phase after loading). A total of eight images are taken to calculate the phase change  $\Delta$ . Before loading, four intensity images with a phase shift increment of  $\pi/2$  are captured, based on Eq. 3, they can be expressed as the following:

$$\begin{aligned} I_1 &= I_0[1 + \gamma \cos(\phi + 0)] \\ I_2 &= I_0[1 + \gamma \cos(\phi + \pi/2)] \\ I_3 &= I_0[1 + \gamma \cos(\phi + \pi)] \\ I_4 &= I_0[1 + \gamma \cos(\phi + 3\pi/2)] \end{aligned} \quad (13)$$



**Fig. 9** The comparison between the (a) intensity fringe pattern and (b) phase map

The random phase  $\phi$  can be calculated using Eq. (14):

$$\phi = \arctan \frac{(I_4 - I_2)}{(I_1 - I_3)} \quad (14)$$

Similarly, the same process can be performed to calculate the random phase after loading using Eq. 4:

$$\phi' = \phi + \Delta = \arctan \frac{(I'_4 - I'_2)}{(I'_1 - I'_3)} \quad (15)$$

where  $I'_1$ ,  $I'_2$ ,  $I'_3$ , and  $I'_4$  are the intensity of 4 images after loading, respectively. Then the phase change  $\Delta$  can be determined by subtracting  $\phi$  from  $\phi'$ :

$$\Delta = \begin{cases} \phi' - \phi; & \text{if } \phi' > \phi \\ \phi' - \phi + 2\pi; & \text{else} \end{cases} \quad (16)$$

#### 4 + 1 Phase Shift Algorithm

Since the 4 + 4 temporal phase shift algorithm requires four images at each loading condition, it is not suitable for dynamic tests (except for harmonic excitation with stroboscopic illumination (Steinchen and Yang 2003)). The 4 + 1 fast temporal phase shift algorithm was proposed to fulfill the dynamic measurement capability (Yang and Siebert 2008). Four images are taken using the same procedure described in the last section before loading. During loading, the camera only records a single image at each load, and no phase shift is needed.

The image captured using the four-step phase shift method before loading can be written using Eq. 13, and the recorded image during loading can be expressed as:

$$I' = I_0[1 + \gamma \cos(\phi + \Delta)] \quad (17)$$

Subtracting Eq. 17 from each equation in Eq. 13, then taking the square of the difference results in the following equations:

$$\begin{aligned}
 \tilde{I}_1^2 &= \{I_0[1 + \gamma \cos(\phi)] - I_0[1 + \gamma \cos(\phi + \Delta)]\}^2 \\
 &= [2I_0\gamma \sin(\phi + \frac{\Delta}{2}) \sin(\frac{\Delta}{2})]^2 \\
 \tilde{I}_2^2 &= \{I_0[1 + \gamma \cos(\phi + \frac{\pi}{2})] - I_0[1 + \gamma \cos(\phi + \Delta)]\}^2 \\
 &= [-2I_0\gamma \cos(\phi + \frac{\Delta}{2} - \frac{\pi}{4}) \sin(\frac{\pi}{4} - \frac{\Delta}{2})]^2 \\
 \tilde{I}_3^2 &= \{I_0[1 + \gamma \cos(\phi + \pi)] - I_0[1 + \gamma \cos(\phi + \Delta)]\}^2 \\
 &= [-2I_0\gamma \cos(\phi + \frac{\Delta}{2}) \cos(\frac{\Delta}{2})]^2 \\
 \tilde{I}_4^2 &= \{I_0[1 + \gamma \cos(\phi + \frac{3\pi}{2})] - I_0[1 + \gamma \cos(\phi + \Delta)]\}^2 \\
 &= [2I_0\gamma \sin(\phi + \frac{\Delta}{2} - \frac{\pi}{4}) \cos(\frac{\pi}{4} - \frac{\Delta}{2})]^2
 \end{aligned} \tag{18}$$

Though  $\Delta/2$  is a low-frequency component (induced by loading), the phase  $\phi$  (phase difference between two points  $P1$  and  $P2$ , see Eq. 3) is the high-frequency component due to surface roughness, and also the terms  $(\phi + \Delta/2)$ , and  $(\phi + \Delta/2 + \pi/4)$  are high-frequency components too. For the high-frequency term, a detector captures an average value of the term within each speckle (Yang and Siebert 2008). With this assumption, we have the following result:

$$\left[ \sin\left(\phi + \frac{\Delta}{2}\right) \right]^2 \approx \frac{1}{2\pi} \int_0^{2\pi} \left[ \sin\left(\phi + \frac{\Delta}{2}\right) \right]^2 d\left(\phi + \frac{\Delta}{2}\right) = \frac{1}{2} \tag{19}$$

Similarly,  $[\sin(\phi + \Delta/2 + \pi/4)]^2$ ,  $[\cos(\phi + \Delta/2)]^2$ , and  $[\cos(\phi + \Delta/2 + \pi/4)]^2$  are equal 1/2 too. Therefore, the Eqs. 18 can be simplified to:

$$\begin{aligned}
 \tilde{I}_1^2 &= 2(I_0\gamma)^2 \sin(\frac{\Delta}{2})^2 = (I_0\gamma)^2 [1 - \cos(\Delta)] \\
 \tilde{I}_2^2 &= 2(I_0\gamma)^2 \sin(\frac{\pi}{4} - \frac{\Delta}{2})^2 = (I_0\gamma)^2 \left[ 1 - \cos\left(\Delta + \frac{\pi}{2}\right) \right] \\
 \tilde{I}_3^2 &= 2(I_0\gamma)^2 \cos(\frac{\Delta}{2})^2 = (I_0\gamma)^2 [1 - \cos(\Delta + \pi)] \\
 \tilde{I}_4^2 &= 2(I_0\gamma)^2 \cos(\frac{\pi}{4} - \frac{\Delta}{2})^2 = (I_0\gamma)^2 \left[ 1 - \cos\left(\Delta + \frac{3\pi}{2}\right) \right]
 \end{aligned} \tag{20}$$

The relative phase change  $\Delta$  can now be solved by:

$$\Delta = \arctan \frac{\left( \tilde{I}_4^2 - \tilde{I}_2^2 \right)}{\left( \tilde{I}_1^2 - \tilde{I}_3^2 \right)} \tag{21}$$

The 4 + 1 phase shift algorithm provides the capability to realize dynamic shearographic measurement using the temporal phase shift technique. However, the phase map quality is not as good as 4 + 4 method due to the assumption in Eq. 19. Therefore, the 4 + 1 algorithm is not widely used in high-accuracy quantitative measurement but can be adopted in some NDT applications.

## Other Temporal Phase Shift Algorithms

Besides the two algorithms mentioned in the previous two sections, there are other temporal phase shift algorithms (Creath 1990; Macy 1983; Wu et al. 2012). Some of them will be briefly introduced in this section.

The first one is the  $3 + 3$  temporal phase shift algorithm. As described in the previous section, three unknowns ( $I_0$ ,  $\gamma$ , and  $\phi$ ) exists in the intensity equation; thus at least three equations are required to solve for the phase  $\phi$ . In the  $3 + 3$  temporal phase shift algorithm, three interferograms with an increment  $2\pi/3$  phase shift are captured at each loading condition, and the phase map  $\phi$  at the loading condition can be calculated. The  $3 + 3$  temporal phase shift algorithm provides similar phase map quality as the  $4 + 4$  temporal phase shift algorithm with fewer captured images. However, the algorithm for calculating phase is not simpler than the  $4 + 4$  algorithm.

Like the  $4 + 1$  fast temporal phase shift algorithm, a  $3 + 1$  fast temporal phase shift algorithm was developed to achieve dynamic measurement using **shearography**. In the  $3 + 1$  temporal phase shift algorithm, three interferograms are taken using the same procedure as the  $3 + 3$  phase shift algorithm before loading, and only one image is required during the loading. Similar to the  $4 + 1$  temporal phase shift algorithm, the phase map quality of the  $3 + 1$  temporal phase shift algorithm is usually adequate for NDT applications.

Other than the  $N + N$  ( $4 + 4$  or  $3 + 3$ ) temporal phase shift algorithms and the  $N + 1$  ( $4 + 1$  or  $3 + 1$ ) fast temporal phase shift algorithms, an  $N + 2$  temporal phase shift algorithm also exists. The  $N + 2$  temporal phase shift uses two phase shifted images during each loading condition to calculate the relative phase change during the test. The phase map quality of these algorithms is better than the  $N + 1$  temporal phase shift algorithms, but they are not as good as  $N + N$  methods. A comparison between these algorithms is shown in Fig. 10.

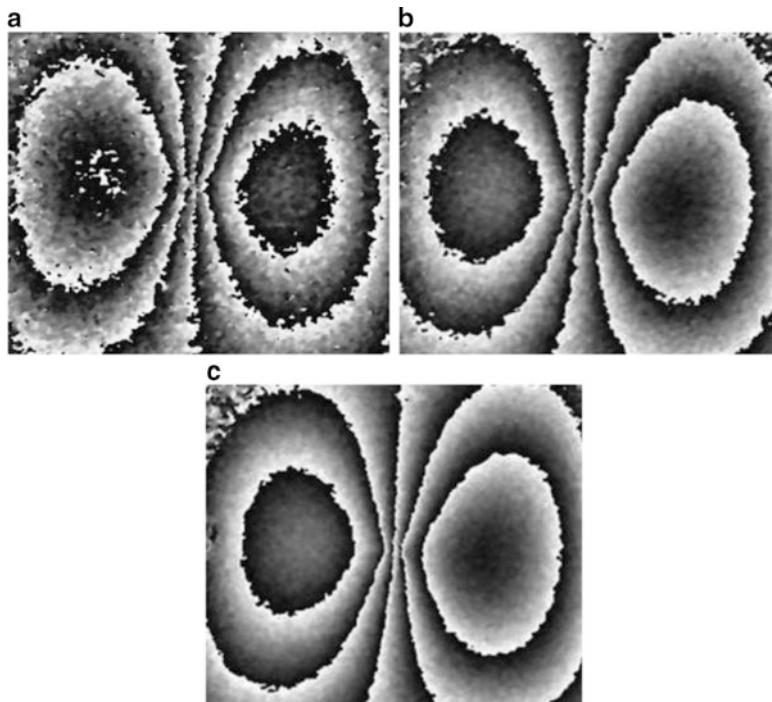
## Spatial Phase Shift Shearography

### Multi-channel Spatial Phase Shift Shearography

#### With a Known Phase Shift

The temporal phase shift technique digital shearography (TPS-DS) achieves a high sensitivity measurement and is very suitable for static tests. Although some of the fast temporal phase shift algorithms can be used for dynamic tests, the quality of the acquired phase map is not good enough for many applications. Spatial phase shift digital shearography (SPS-DS) was proposed to overcome this drawback. Spatial phase shift digital shearography uses only a single image to create the phase map, making this technique extremely suitable for dynamic tests. The SPS-DS can be categorized into two types: multi-channel spatial phase shift shearography and carrier-frequency spatial phase shift shearography.

The multi-channel approach determines the phase distribution using the intensity data of three adjacent pixels in a single interferogram (Yamaguchi 2006;



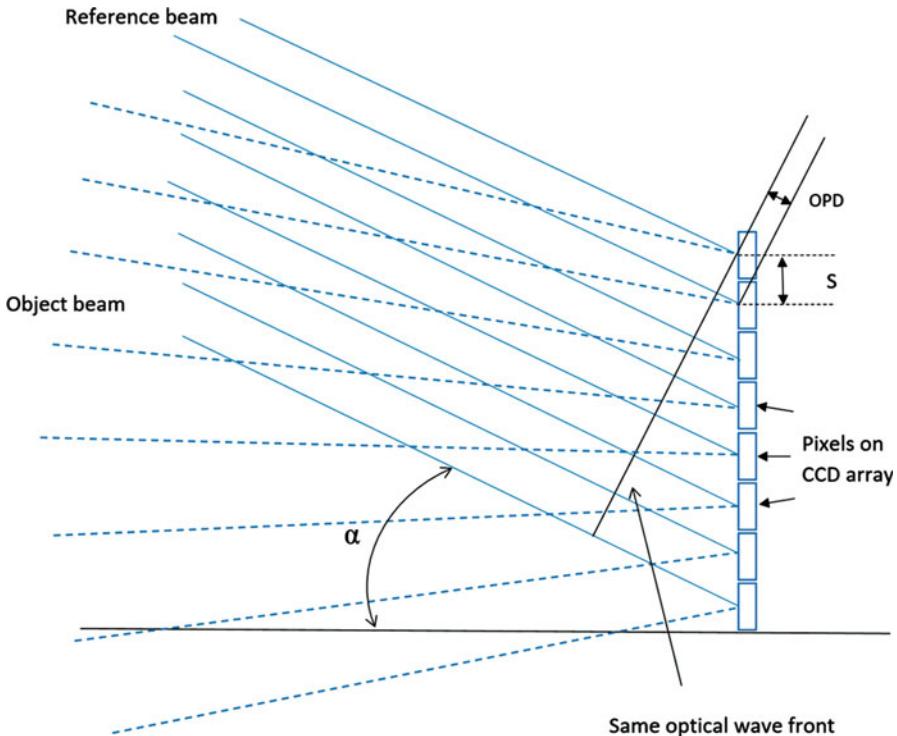
**Fig. 10** Phase Map obtained by (a) 3 + 1 fast algorithm; (b) 3 + 2 algorithm; (c) 3 + 3 algorithm

Hettwer et al. 2000). Figure 11 shows a schematic of this approach. If the phase difference  $\delta$  between the pixels is fixed, the equations described in section “[4 + 4 Phase Shift Algorithm](#)” with four equations or with three equations can be used to solve the phase map using the single interferogram. If  $\delta = 90^\circ$ , with the above arrangement, the intensity of four adjacent pixels can be described using the following equations:

$$\begin{aligned} I_1 &= I_0[1 + \gamma \cos(\phi)] \\ I_2 &= I_0[1 + \gamma \cos(\phi + \delta)] \\ I_3 &= I_0[1 + \gamma \cos(\phi + 2\delta)] \\ I_4 &= I_0[1 + \gamma \cos(\phi + 3\delta)] \end{aligned} \quad (22)$$

The random phase  $\phi$  can be calculated by the Eq. (14) used in the four-step temporal phase shift algorithm.

The introduced phase difference  $\delta$  can be calculated based on the relation shown in Eqs. 23. The optical path difference between adjacent pixels  $\Delta L$  can be controlled by setting the reference beam at an appropriate angle  $\alpha$ :  $\Delta L = S \cdot \sin \alpha$ , where  $S$  is the pixel size. The relationship between the phase difference and the angle is given:



**Fig. 11** Schematic of multi-channel approach spatial phase shift technique

$$\delta = 2\pi \cdot \frac{\Delta L}{\lambda} = 2\pi \cdot \left( S \cdot \frac{\sin \alpha}{\lambda} \right) \quad (23)$$

where  $\lambda$  is the laser wavelength. For given wavelength and pixel size, the phase difference  $\delta$  can be determined by setting an appropriate reference beam angle  $\alpha$ . After the object is loaded, the same procedure can be used to calculate the phase distribution  $\phi'$  at the loaded condition.

### With an Unknown Phase Shift

The multi-channel SPS-DS algorithm described in section “[With a Known Phase Shift](#)” by using 3 or 4 pixels method requires a very accurately known incident angle  $\alpha$ . As a result, the measurement result of the described method is limited to the accuracy of the incident angle  $\alpha$ . Because of the difficulty to set a known incident angle, a more flexible multi-channel SPS-DS algorithm was proposed to solve this problem (Lassahn et al. 1994). In this method, the same experimental setup is used to induce the incident angle. The difference is that the incident angle is regarded as an unknown constant rather than the known constant. In this case, since the introduced phase difference  $\delta$  becomes an unknown, at least four equations are required to solve the random phase  $\phi$ . A well-known five-step phase shift algorithm is used for this

case. The equations are described in the following: assuming the introduced phase difference between pixels is constant, the intensity of the five adjacent pixels can be expressed as:

$$\begin{aligned} I_1 &= I_0[1 + \gamma \cos(\phi)] \\ I_2 &= I_0[1 + \gamma \cos(\phi + \delta)] \\ I_3 &= I_0[1 + \gamma \cos(\phi + 2\delta)] \\ I_4 &= I_0[1 + \gamma \cos(\phi + 3\delta)] \\ I_5 &= I_0[1 + \gamma \cos(\phi + 4\delta)] \end{aligned} \quad (24)$$

The random phase  $\phi$  can be calculated using the following equation:

$$\phi = \arctan \left\{ \frac{2(I_2 - I_4)}{2I_3 - I_1 - I_5} \cdot \frac{1}{\sqrt{1 - \left[ \frac{I_1 - I_5}{2(I_2 - I_4)} \right]^2}} \right\} \quad (25)$$

It should be emphasized that the background ( $I_0$ ), the contrast ( $\gamma$ ), and the phase ( $\phi$ ) of the intensity equations at these adjacent pixels must keep identical. This is the necessary condition for the multi-channel spatial phase shift methods. To maintain this necessary condition, the speckle size should be controlled so that it can cover 3, or 4 or 5 adjacent pixels depending on the algorithms utilized. This can be achieved by adjusting the aperture of the camera used. However, too big speckle size (to cover more pixels) needs very small aperture which will result in need of high power laser and low measuring resolution.

## Carrier-Frequency Spatial Phase Shift Shearography

### Method Based on Mach-Zehnder Interferometer

The carrier-frequency technique is widely used in many applications, such as in the telecommunication area. The carrier signal is modulated by a high-frequency carrier, and this technique can also be used to solve the phase from a single interferogram (Bhaduri et al. 2006, 2007; Xie et al. 2013a, b). A typical carrier-frequency spatial phase shift shearography can be explained by the Mach-Zehnder interferometer-based shearography setup. This setup has been shown in Fig. 4. The carrier frequency is introduced by tilting one mirror in the Mach-Zehnder interferometer.

After passing the shearing device, the light is divided into a sheared part and an un-sheared part. The wavefront can be expressed as:

$$\begin{aligned} u_1(x,y) &= |u_1(x,y)| \exp[i\phi(x,y)] \\ u_2(x,y) &= |u_2(x + \Delta x, y + \Delta y)| \exp[i\phi(x + \Delta x, y + \Delta y) - 2\pi i \cdot f_c \cdot x] \end{aligned} \quad (26)$$

Where  $u_1$  and  $u_2$  are the two components with and without shearing.  $\Delta x$  and  $\Delta y$  represent the shearing distance in the  $x$  and  $y$ -direction, respectively. The carrier frequency  $f_c$ , which is introduced by the tilting angle  $\beta$ , can be expressed as:

$$f_c = \frac{\sin \beta}{\lambda} \quad (27)$$

Thus, the intensity recorded by the digital camera can be represented as:

$$\begin{aligned} I &= (u_1 + u_2)(u_1^* + u_2^*) \\ &= u_1 u_1^* + u_1 u_2^* + u_2 u_1^* + u_2 u_2^* \end{aligned} \quad (28)$$

Where the  $*$  denotes the complex conjugate of the light beams. To extract the phase  $\phi$  from the above equations, a Fourier transform is performed, and the recorded information can be expressed in the frequency domain using the Fourier transform:

$$\begin{aligned} F(I) &= U_1(f_x, f_y) \otimes U_1^*(f_x, f_y) \\ &\quad + U_2(f_x + f_c, f_y) \otimes U_2^*(f_x + f_c, f_y) \\ &\quad + U_1(f_x, f_y) \otimes U_2^*(f_x + f_c, f_y) \\ &\quad + U_2(f_x + f_c, f_y) \otimes U_1^*(f_x, f_y) \end{aligned} \quad (29)$$

Where  $\otimes$  is the convolution operation. Three spectra can be observed in the frequency domain, as shown in Fig. 12. The center spectrum is the low-frequency term  $F(u_1 u_1^* + u_1 u_2^*)$ , and contains the information from the background. The two side spectra are the terms with the carrier frequency  $F(u_2 u_1^*)$  and  $F(u_1 u_2^*)$ . These two terms are shifted to the location  $(f_c, 0)$  and  $(-f_c, 0)$ . The spectral width is determined by the size of the aperture (AP).

The phase can be extracted by applying a windowed inverse Fourier transform on the area located at  $(-f_c, 0)$ , and it can be expressed as:

$$\phi + 2\pi \cdot f_c \cdot x = \arctan \frac{\text{Im}[F(u_1 u_2^*)]}{\text{Re}[F(u_1 u_2^*)]} \quad (30)$$

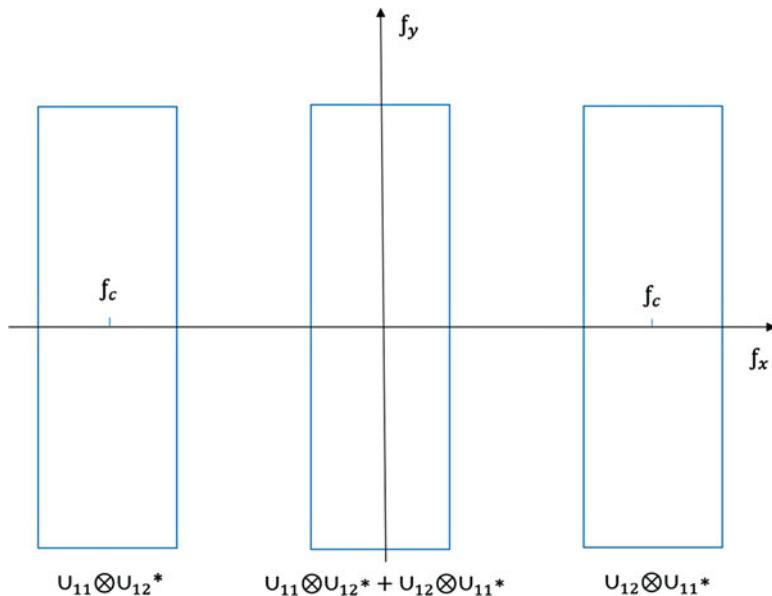
Where the  $\text{Im}$  and  $\text{Re}$  denote the imaginary and real component of the complex number. The phase after loading can be calculated using the same method:

$$\phi' + 2\pi \cdot f_c \cdot x = \arctan \frac{\text{Im}[F(u'_1 u'_2^*)]}{\text{Re}[F(u'_1 u'_2^*)]} \quad (31)$$

The relative phase change due to the loading can be calculated as:

$$\Delta = \phi' + 2\pi \cdot f_c \cdot x - (\phi + 2\pi \cdot f_c \cdot x) = \phi' - \phi \quad (32)$$

In this way, the phase map is evaluated using a single image at each loading condition, and the acquisition rate is no longer limited by the phase shift procedure.



**Fig. 12** Schematic of the frequency domain using the carrier frequency spatial phase shift technique

### Method Based on Michelson Interferometer

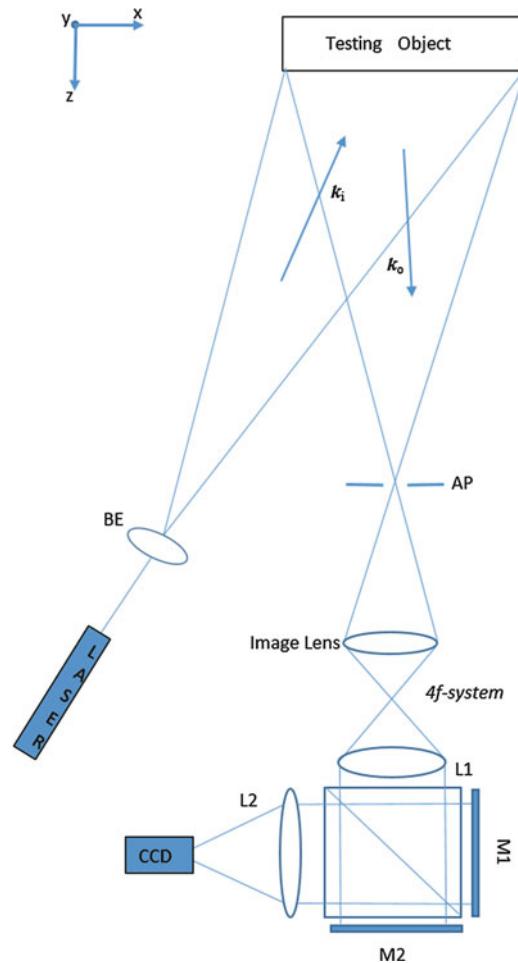
Another popular carrier-frequency spatial phase shift digital shearography setup is based on the 4f-system embedded Michelson interferometer. This setup is shown in Fig. 13. The fundamental of this setup is similar to the conventional modified Michelson interferometer-based digital shearography setup, which is described in the previous section. The 4f-system embedded Michelson interferometer-based SPS-DS has the following changes: (1) a 4f-system is embedded; and (2) an aperture is placed in front of the imaging lens. The 4f-system allows the field of view of the measurement system to be adjusted and enlarged, and the aperture is used to limit the spatial frequency of the light. The procedure of the phase evaluation of this setup is the same as the carrier-frequency SPS-DS based on Mach-Zehnder Interferometer.

The Michelson interferometer-based spatial phase shift shearography has a simpler setup than Mach-Zehnder Interferometer based spatial phase shift shearography. However, the Michelson interferometer-based system usually needs a large shearing amount so that the two side spectra can be separated from the central spectrum whereas the Mach-Zehnder based setup can do it easily either in small or large shearing amount.

### Comparison and Discussion

Spatial phase shift digital shearography (SPS-DS) was proposed to achieve dynamic measurement. SPS-DS can be categorized into two types: multi-channel spatial phase shift shearography and the carrier-frequency spatial phase shift shearography. The multi-channel approach determines the phase distribution using the intensity data

**Fig. 13** Schematic of the 4f-system embedded carrier frequency spatial phase shift shearographic setup



of three or four adjacent pixels in a single interferogram; while the carrier-frequency approach extracts the phase distribution from the frequency domain via Fourier transform. The algorithm complexity of the multi-channel approach is much lower than that of the carrier-frequency approach. However, the phase map quality of the carrier-frequency approach is much better than the multi-channel approach.

## Practical Applications of Digital Shearography for NDT

### Introduction

Shearography, as a laser-based optical method, has been widely used in the automotive and aerospace industries for NDT. Its advantages include its high sensitivity, robustness for rigid-body motion, and high measurement accuracy

with the phase shift technique. However, a successful application of digital shearography for NDT still depends on many factors, including the depth and type of defect, the type of material, the correct shearing amount and direction, and the illumination and loading. In this section, the effects of these parameters will be analyzed and discussed.

## Instrumentation

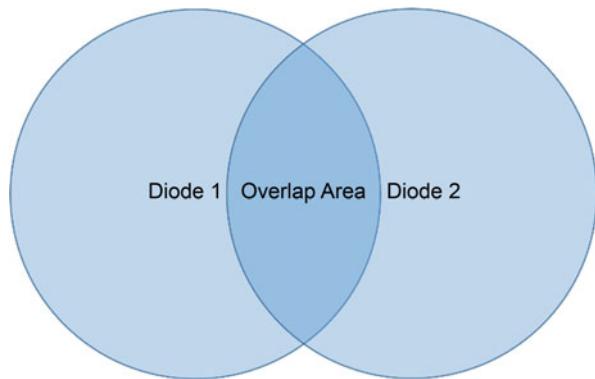
The schematic setup has been described in detail in the previous sections; this section will summarize the hardware of digital shearography. The hardware for digital shearography consists of three parts: a digital camera, a shearing unit, and a laser device for illumination.

The digital camera records intensity through the use of gray levels. Various digital cameras are available commercially. The main specification of a digital camera is sensor type, sensor size, resolution, and frame rate (Li et al. 2017a and b). The sensor of the digital camera can be categorized into two types: complementary metal-oxide semiconductor (CMOS) and charge-coupled device (CCD). CCD cameras used to dominate the area of industrial application; however, the CMOS camera has started to become the more favorable type due to the high development of the CMOS technique. The sensor size is another important parameter for the digital camera. Sensor sizes commonly used for the industrial camera range from half an inch to 1 inch. A larger sensor size camera always provides better digital image quality. Corresponding to the sensor size, another important parameter of the digital camera is the resolution. Currently, the resolution of the most common industrial cameras is between 2 megapixels and 8 megapixels. A higher resolution camera can record a more detailed image and provide a more detailed phase map for shearography. The last important parameter is the frame rate. The frame rate of the digital camera determines the upper limit of the dynamic range of the test system. The frame rate for the current standard industrial camera is between 10 Hz and 100 Hz. If the dynamic range of the test is high, a high-speed digital camera is required.

The shearing device is the key component of digital shearography. It brings lights scattered from two points on the test object surface to a single point on the image plane. Different devices can be used as the shearing device, such as an optical glass wedge (Hung 1982), a bi-angle prism (Ettmeyer 1991), a double-refractive prism (Hung et al. 2003), a modified Michelson interferometer (Yang et al. 1995), and a modified Mach-Zehnder interferometer (Pedrini et al. 1996). Furthermore, it is important to select a shearing device that is compatible with the phase shift technique being utilized. The methodology of the different shearing devices and the phase shift technique has been introduced in detail in the previous section and will not be repeated here.

The laser device for illumination is the third important part of digital shearography in the aspect of hardware. A strong laser source is required to perform a whole-field test. Traditional high-powered laser sources, such as a He-Ne laser and Nd:YAG laser, are too big, fragile, and expensive for the NDT applications. Based on

**Fig. 14** Illustration of the shearographic system with two diodes



the fundamentals of shearography, the interference phenomenon in shearography is formed by two waves from two different points, and the optical path difference between them is very small. Therefore, the coherence length requirement for the laser device is relaxed. This characteristic of shearography allows the laser diode to be used as the laser source for illumination. The laser diode does not have a long coherent length, but it is compact in size and cheap in price. A portable digital shearography setup can be built utilizing of the laser diode in the shearography system. Also, multiple laser diodes can be used at the same time for large area illumination in some NDT applications, as shown in Fig. 14 (Yang 1998; Kalms and Osten 2003; Jackson 2004).

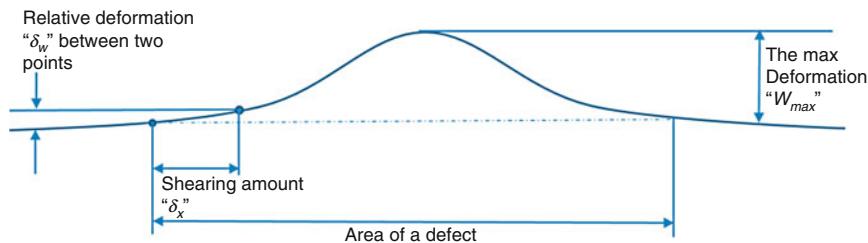
## Measuring Sensitivity

### Sensitivity for Phase Measurement

As displayed by Eqs. 11 and 12, digital shearography measures the gradient of deformation by determining the relative phase change. Therefore, the sensitivity of the shearography system depends on the sensitivity of the phase measurement. The measuring sensitivity of the phase for the shearography without using the phase shift technique is around  $2\pi$ . However, the measuring sensitivity can be increased dramatically using the phase shift technique. Theoretically, the phase shift technique can reach a sensitivity of  $2\pi/256$  for the phase measurement, if the depth resolution of the camera is 8 bits. Practically, the sensitivity of the phase measurement is from  $2\pi/30$  to  $2\pi/10$ , depending on the level of the speckle noises and the application of the phase shift technique.

### Sensitivity for Shearing Amount

Another parameter that impacts the measuring sensitivity of shearography is the shearing amount. The gradient of deformation is obtained by calculating the deformation difference between two points with a separation of the shearing amount. Obviously, the measuring sensitivity depends on the shearing amount (Steinchen et al. 1998).



**Fig. 15** The relation among the deformation, the relative deformation, and the shearing amount at the defect location

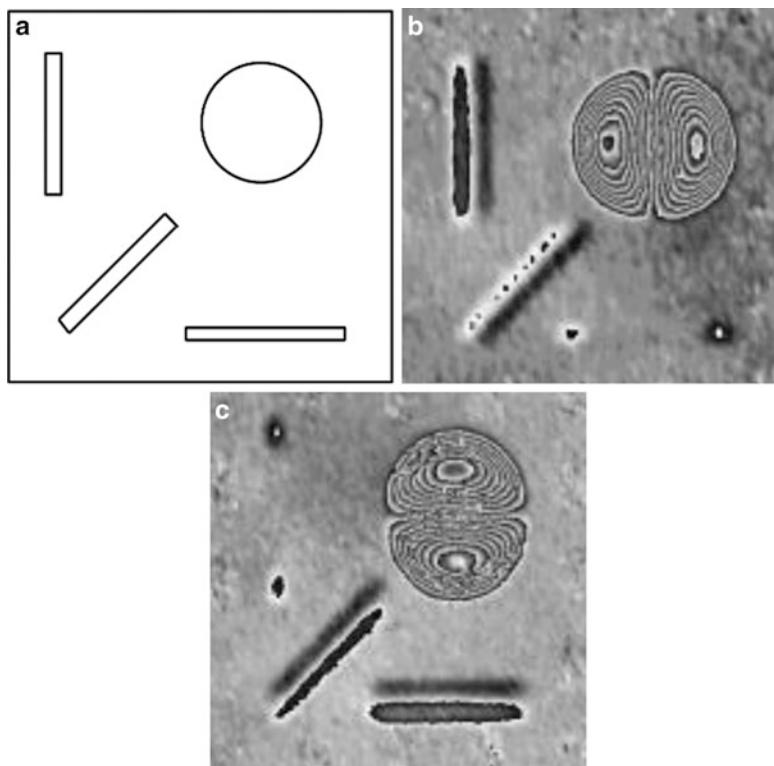
A small shearing amount leads to a low measuring sensitivity since the relative deformation between two points with small separation is low. However, the measuring result is close to the first derivative. For strain measurement, a small shearing amount is desirable. However, in NDT for measuring small defects, a relatively large shearing amount is required to reach a high measuring sensitivity.

Thus, in one respect, a small shearing amount is favorable to obtain the accurate strain information; on the other hand, a large shearing amount leads to high measuring sensitivity. Therefore, a critical shearing amount exists which provides the maximum sensitivity of the system. If the shearing amount is larger than the critical shearing amount, the sensitivity of the system will not increase anymore. To explain this concept, a relationship among deformation, relative deformation, and the shearing amount over the defect area is shown in Fig. 15. The relative deformation between two points increases with the shearing amount; however, the relative deformation no longer increases after the shearing amount reaches one-half of the area of deformation. Therefore, the critical shearing amount should be equal to one-half of the defect diameter. Usually, one knows what is the smallest defect size to be inspected, the critical shearing amount should be the one-half of the smallest defect size.

### Sensitivity for Shearing Direction

The shearing direction can also affect the sensitivity of shearography. If the defect is a circular form, the shearing direction won't affect the sensitivity of the measurement result since the anomaly of the relative deformation shows a uniform appearance in all directions. However, if the defect is a narrow slot form, the shearing direction plays an important role in flaw detection because the anomaly of the relative deformation is sensitive to the direction. To demonstrate this effect, a square plate with one round groove and three narrow slits oriented in different directions on the plate's back side, as shown in Fig. 16, is tested with different shearing directions (x and y-direction). The slit oriented in the horizontal direction cannot be detected properly using the horizontal shearing direction, whereas the vertical notch can only be detected using the horizontal shearing direction.

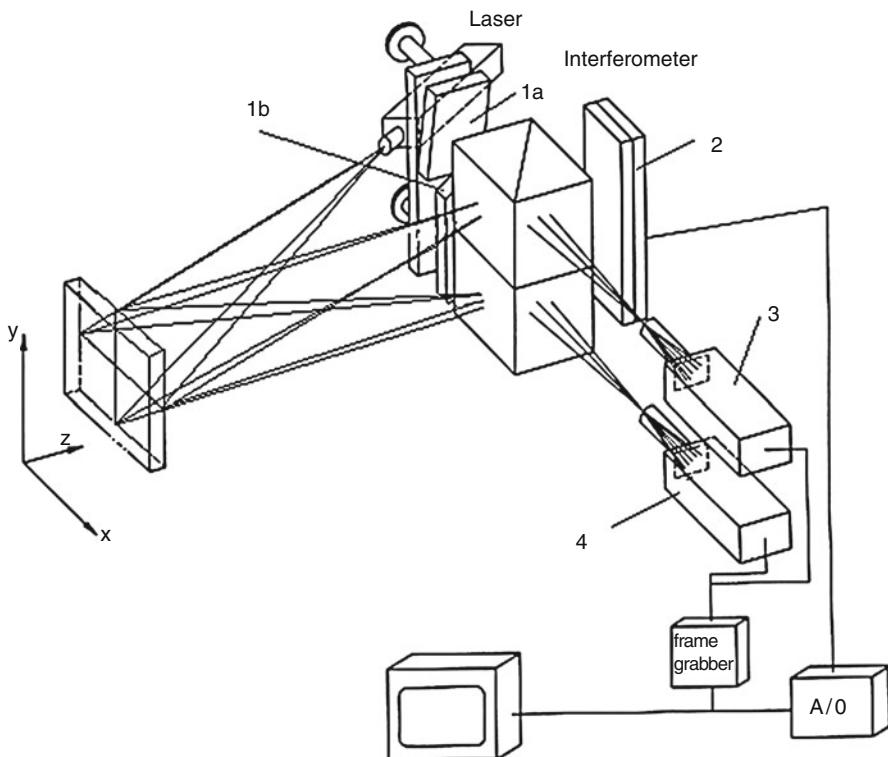
Shearographic setup with double shearing direction has been developed for both the temporal phase shift (TPS) technique and spatial phase shift (SPS)



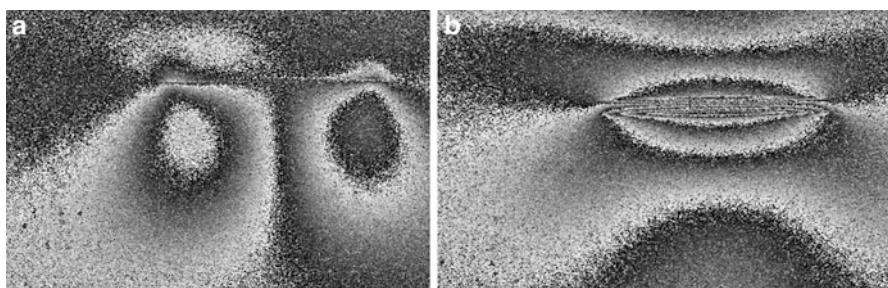
**Fig. 16** NDT result with different shearing direction. (a) Location and the shape of the defect (left). (b) NDT result with the x-direction shearing (middle). (c) NDT result with the y-direction shearing

technique to solve the above problem. A TPS technique based dual-shearing direction shearography system using two modified Michelson interferometers has been reported in the literature (Yang et al. 2004). The schematic of the system is shown in Fig. 17. Two beam splitters are stacked on top of each other. The two mirrors (1a and 1b) are arranged on one side of the beam splitters. On the other side of the beam splitters, a PZT driven mirror (2) is used to induce the temporal phase shift. Two cameras (3 and 4) are also stacked on top of each other and correspond to the two beam splitters. The different shearing directions for the upper and lower system are produced by tilting the upper (1a) and lower (1b) mirrors in different directions. Thus, two interferograms can be taken with the different shearing directions, and the dual-shearing direction TPS-DS is achieved.

Figure 18 shows the TPS technique based shearographic system with two shearing directions for NDT of a micro-crack in a glass fiber-reinforced composite tube with loading by internal pressure. Two shearograms were obtained simultaneously under the same loading. The micro-crack was demonstrated more clearly in the shearogram with y-shearing direction.

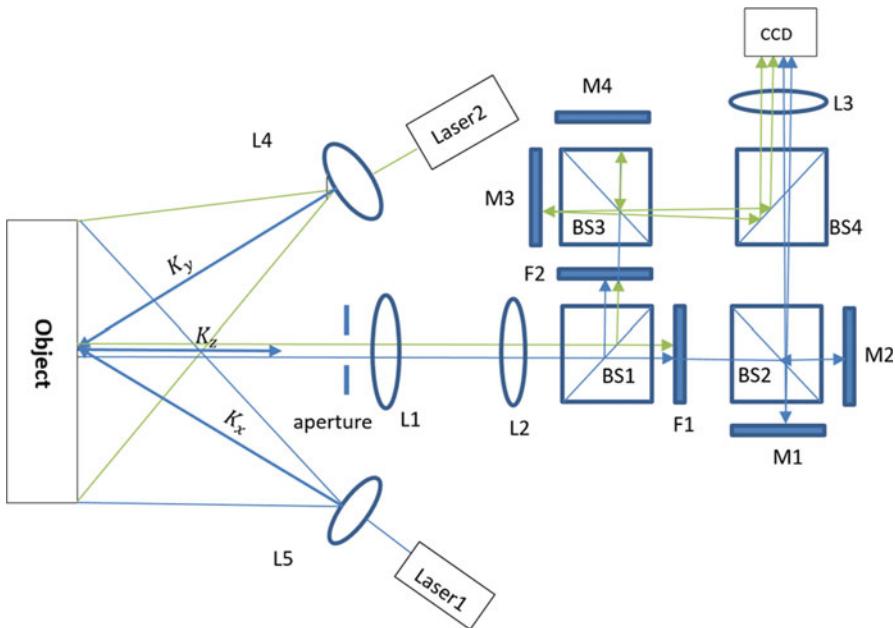


**Fig. 17** A TPS technique based shearography with two shearing directions



**Fig. 18** TPS based shearography with two shearing directions for NDT of a micro-crack: (a) x-shearing, (b) y-shearing

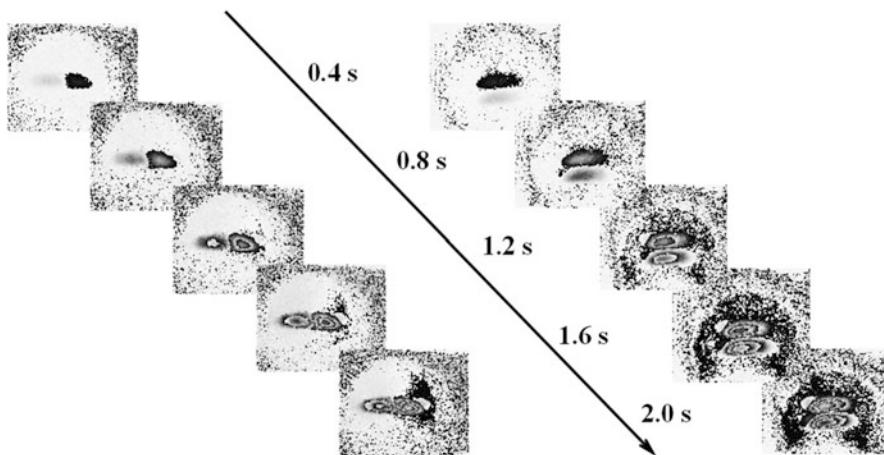
The spatial phase shift technique based dual-shearing direction shearography has been developed recently. Unlike the temporal phase shift technique based system, only one camera is required using the carrier-frequency technique. The schematic of the system is shown in Fig. 19. Two lasers with different wavelengths are used to illuminate the object. The scattered light from the object passes the shearing device



**Fig. 19** Schematic of the dual-shearing direction spatial phase shift shearography

and forms the image on the digital camera. The shearing device is composed of two modified Michelson interferometers: one generates shearing in the vertical direction, and the other generates shearing in the horizontal direction. The carrier frequency is also generated at the same time during shearing. Due to the different shearing directions, the carrier frequency of the two interferometers is different. Thus, the phase map from the two interferometers can be separated in the frequency domain. Two band filters, corresponding to the wavelengths of the two lasers, are used to guarantee that only one laser wave passes to one modified Michelson interferometer. An aperture is placed in front of the 4f imaging system to reduce the spatial frequency of the wave, and a beam splitter is used to synthesize two beams from these two Michelson interferometers to a single digital camera. The details regarding the principle of developing spatial phase shift technique based dual-shearing direction shearography systems can be found in the literature (Xie et al. 2013a, b, 2015; Wang et al. 2016).

Figure 20 shows SPS technique based shearographic system with two shearing directions for NDT of a honeycomb structure under a continuous loading. A heat gun was used to load the honeycomb continuously for 2 s, which resulted in a temperature change on the surface. At the same time, the SPS technique based shearographic system captured the speckle pattern images at a frame rate of 10 fps. Then, the phase map history in the  $x$  and  $y$  directions during the 2 s was measured simultaneously. Because the SPS based shearography is particularly well suited for dynamic loading, the shearogram history along the 2 s loading, i.e., the growth of



**Fig. 20** SPS based shearography for NDT of a honeycomb structure under a continuous loading in two shearing directions simultaneously, continuously, and dynamically

$\partial w / \partial x$  (x-shearing) and  $\partial w / \partial y$  (y-direction) during the 2-s loading, can be clearly observed. This NDT example shows that with SPS technique based shearographic system is possible to do NDT measurements in dual-shearing directions simultaneously, continuously, and dynamically.

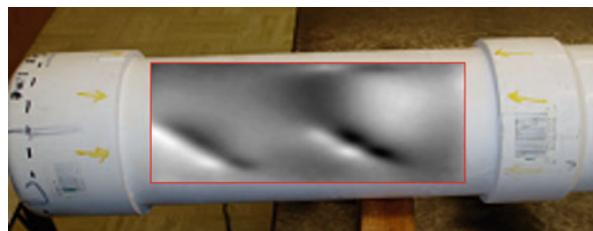
## Methods of Stressing

Various methods are currently used to induce the stress for NDT (Tesfamariam and Goda 2013). Though shearography is relatively insensitive to rigid-body motion, the stressing methods should not create intolerable (too large) rigid-body motion of the test object, but generate enough stress to detect the flaws. Internal pressure loading, vacuum loading, thermal loading, and dynamic loading are most common methods used to induce stress.

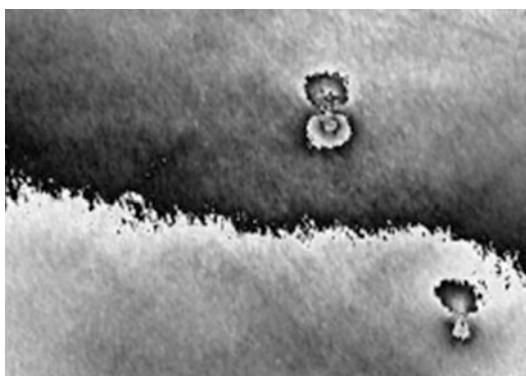
Pressure vessels, pipes, or other structures that can be pressurized internally can be loaded using the internal pressure. The internal pressure can be applied to the maximum value of the actual loading condition to represent actual stressing of the structure. Shearography would reveal the flaws that create strain concentrations. The revealed critical flaws would reduce the strength of the structure and lead to failure of the component. As an example, a plastic PVC pipe with two flaws is loaded by a 0.02 MPa internal pressure. The NDT result using shearography, overlapped on the object, is shown in Fig. 21. The stress concentration can be clearly detected by viewing the phase map.

Vacuum stressing is one of the most popular loading methods for revealing disbands and delaminations in the composite material. Vacuum stressing is equivalent to a uniform tensile force applied to the surface, pulling the surface outward. The flaws under the vacuum stressing would bulge out slightly, and the anomalies

**Fig. 21** NDT result of an internal pressured pipe using shearography



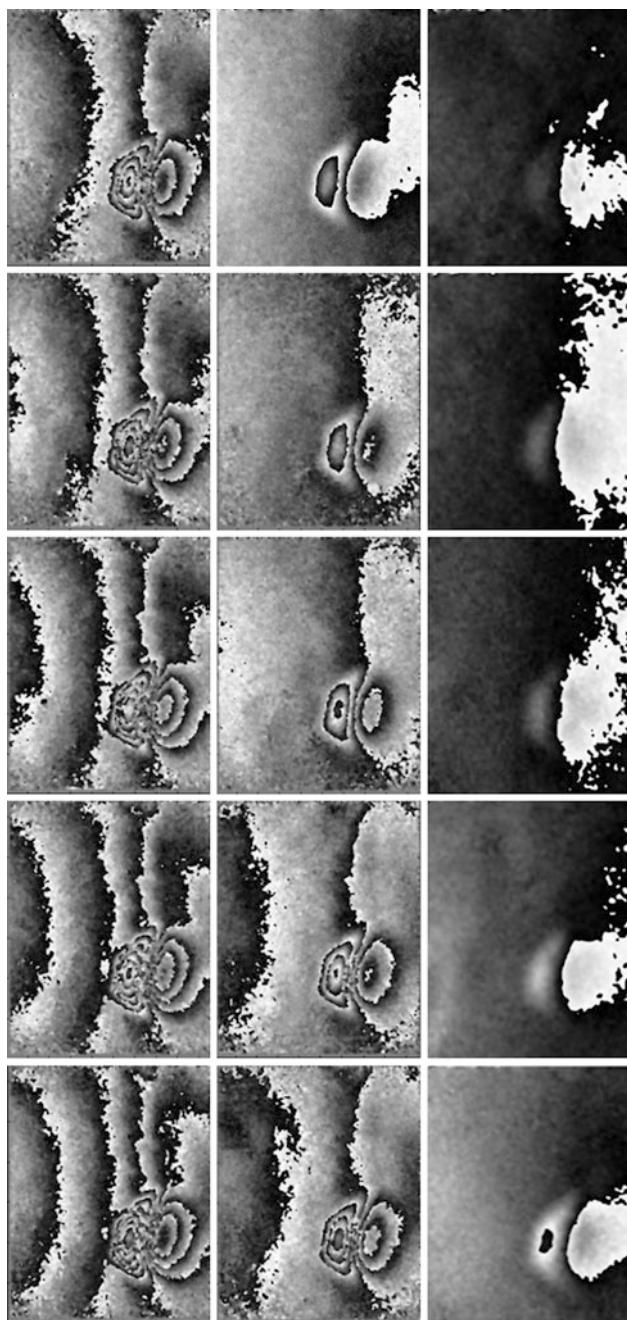
**Fig. 22** NDT result of the delamination on the fuselage under vacuum stressing



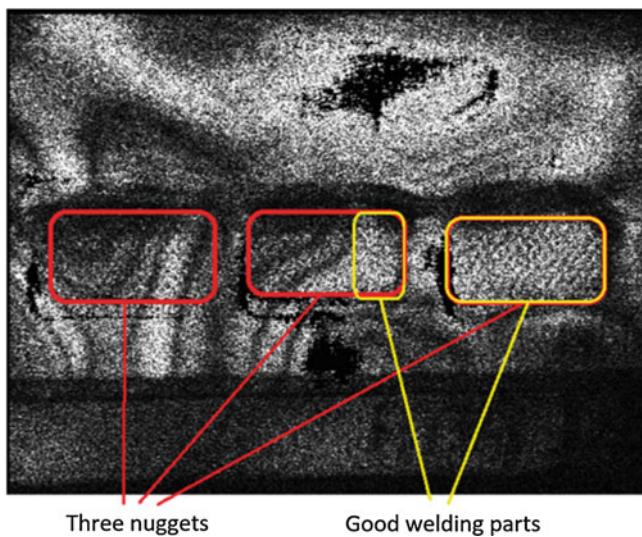
due to the bulge can be detected using digital shearography. Figure 22 shows the NDT result for a composite fuselage using the vacuum stressing method. The shearing direction is in the vertical direction, and the two de-laminations are clearly demonstrated.

Thermal stressing, another popular loading method, is an alternative method to generate the stress. The developed temperature gradient, generated by the thermal stressing, induces thermal strain to the object. Heat causes disbonds or delaminations expand and causes strain anomalies on the surface. Figure 23 shows NDT results using  $3 + 1$  temporal phase shift shearography for real-time measurement under thermal loading. The sample, an aerospace honeycomb aluminum plate with delamination, was pre-heated to a certain temperature. A series of interferograms are taken during cooling of the plate using the procedure that was described in the previous section.

Dynamic stressing is also useful loading method in applications for inspecting bonding integrity and weld joints. Dynamic stressing is created by an excitation source. The excitation source can be divided into harmonic and non-harmonic excitation sources. The harmonic excitation is usually generated by the PZT or a shaker and is suitable for dynamic analysis. The non-harmonic excitation is typically generated by an impact loading, and it usually uses the double-exposure method, or if phase map is required, it should use the double-exposure SPS-shearography system. Figure 24 shows an example for NDT of welding condition of nuggets in an automotive battery. While the top surface of the battery is harmonically excited at 11 KHz, the nugget area where multi sheets are welded together starts to vibration, a



**Fig. 23** NDT result of the delamination on the honeycomb structure plate under heating using the  $3 + 1$  fast temporal phase shift shearography



**Fig. 24** NDT result of the welding parts under dynamic stressing

fringe pattern can be observed in the nugget areas (the left and the middle nuggets), whereas no fringes can be observed in the good nugget area (right nugget). It indicated that the left two nuggets are detached since the fringe characteristic on the nuggets under vibration are similar to the free copper margin area. However, there are no fringes on the right nugget, which means it is perfectly welded.

---

## Conclusions

### Potentials

In comparison with other optical methods for NDT, shearography offers many advantages, such as:

1. Shearography directly measures the first derivative of deformation/displacement which is the strain information, defects in objects generate strain concentration. Thus, shearography reveals defects in an object by identifying strain anomalies through the display of strain concentrations, which is a more direct and clear way to reveal defects as shown in Fig. 8b.
2. Shearography is relatively insensitive to rigid-body movement because a rigid-body movement generates a displacement, but not the first derivative of displacement. Therefore, shearography is less sensitive to environmental disturbances and suited well for practical application in compared with other interferometric methods such as holography, electron speckle pattern interferometry, etc.

3. Shearography uses a self-reference interferometric system, which has a very simple optical layout. It is of benefit to building a very robust shearographic sensor for industrial applications.
4. The self-reference interferometric system has balanced optical paths. Thus, a high requirement of coherent length is greatly reduced, which enables the utilization of a cost-economical diode-laser for illumination. If a larger object surface is needed for inspection, multi-diode lasers can be applied.

In comparison with the early version of shearography (so-called electronic shearography), in which the intensity image is used as the measured result, the phase shift digital shearography offers many new possibilities for NDT, such as:

1. The measurement sensitivity of the temporal phase shift digital shearography for phase determination is much higher than the electronic version. The sensitivity of the electronic shearography for phase determination is  $2\pi$ , however, the sensitivity of the digital shearography with phase shift technique could reach  $2\pi/20$  or higher, which enables inspection of smaller defect as shown in Fig. 9b.
2. Phase shift digital shearography with multi-shearing directions makes NDT of direction sensitive flaws, such as micro-cracks, narrow slot shape defects, easier.
3. Spatial phase shift digital shearography enables displaying shearograms under dynamic loading continuously and dynamically.

## Limitations

Digital shearography is an experimental technique for surface measurement. Thus, some limitations still exist for NDT.

1. Shearography detects flaws by looking for anomalies of the surface under stress. However, flaws that lie far from the object surface are hard to detect, since the flaw cannot generate large enough surface anomalies to be detected by shearography.
2. The application of shearography for NDT is mainly limited to composite materials, honeycomb structures, and thin plates. High strength materials, such as metallic materials, are difficult to test using shearography.
3. Although shearography uses a “self-referencing” optical system and is relatively robust to environmental disturbance compared to the other laser technique, shearography is still sensitive to ambient noises such as large rigid-body motions and strong thermal airflow, because it is, after all, an interferometric method.
4. With the fast development of the digital technique, the resolution of the digital camera has increased dramatically. However, the spatial resolution of the digital camera is still much lower than the special film used in photographic shearography.
5. The phase map obtained by the phase shift technique still contains lots of noise, and the signal to noise ratio of current shearography, especially, in the spatial phase shift digital shearography, is not high enough to detect some minor flaws.

## Summary

The fundamental of shearography for NDT and its applications have been presented in this chapter. The principles of digital shearography including optics, setups, and algorithms are detailed described and analyzed in the first part of the chapter. Phase-shift techniques including temporal and spatial phase-shift methods are then introduced which can be applied to improve the measuring sensitivity of shearography for NDT. Different NDT applications of shearography are discussed and investigated in the last part, along with the analysis of the measuring sensitivity, and the concept of the critical shearing amount etc. In addition, higher dynamic test range can be achieved using the recently developed spatial phase-shift digital shearography. Digital shearography is well suitable to detect the delamination, disbonds, and microcracks in composite materials, honeycomb structures, and thin plates. It is expected that the digital shearography can be applied to a wider range of applications with the recent and future developments.

---

## References

- ASTM E2581-14 (2014) Standard practice for Shearography of polymer matrix composites and sandwich Core materials in aerospace applications. ASTM Int, West Conshohocken
- Baldwin JM, Bauer DR (2008) Rubber oxidation and tire aging-a review. *Rubber Chem Technol* 81(2):338–358
- Bhaduri B, Mohan NK, Kothiyal MP, Sirohi RS (2006) Use of spatial phase shifting technique in digital speckle pattern interferometry (DSPI) and digital shearography (DS). *Opt Express* 14(24):11598–11607
- Bhaduri B, Mohan NK, Kothiyal MP (2007) Simultaneous measurement of out-of-plane displacement and slope using a multiaperture DSPI system and fast Fourier transform. *Appl Opt* 46(23):5680–5686
- Burleigh DD (2002) Portable combined thermography/shearography NDT system for inspecting large composite structures. In: Thermosense XXIV, vol 4710. International Society for Optics and Photonics, the International Society for Optical Engineering, Bellingham, USA, pp 578–588
- Cartz L (1995) Nondestructive testing. ASM International. The Materials Information Society, Materials Park
- Clark MR, McCann DM, Forde MC (2003) Application of infrared thermography to the non-destructive testing of concrete and masonry bridges. *NDT E Int* 36(4):265–275
- Creath K (1990) Phase-measurement techniques for nondestructive testing. In: Proceedings of SEM conference on hologram interferometry and speckle metrology, Baltimore, pp 473–478
- De Angelis G, Meo M, Almond DP, Pickering SG, Angioni SL (2012) A new technique to detect defect size and depth in composite structures using digital shearography and unconstrained optimization. *NDT E Int* 45(1):91–96
- Diaz FV, Kaufmann GH, Galizzi GE (2000) Determination of residual stresses using hole drilling and digital speckle pattern interferometry with automated data analysis. *Opt Lasers Eng* 33(1):39–48
- Etemeyer A (1991) Shearografie-ein optisches verfahren zur zer-stoerungsfreien werkstoff-pruefung. *Tm-Technisches Messen* 58:247
- Groves RM, James SW, Tatam RP (2004) Shape and slope measurement by source displacement in shearography. *Opt Lasers Eng* 41(4):621–634

- Gryzgoridis J, Findeis D (2008) Benchmarking shearographic NDT for composites. *Insight-Non-Destr Test Cond Monit* 50(5):249–252
- Hathaway RB, Hovanesian JD, Hung MYY (1997) Residual stress evaluation using shearography with large-shear displacements. *Opt Lasers Eng* 27(1):43–60
- Hettwer A, Kranz J, Schwider J (2000) Three channel phase-shifting interferometer using polarization-optics and a diffraction grating. *Opt Eng* 39(4):960–967
- Huang YH, Ng SP, Liu L, Li CL, Chen YS, Hung YY (2009) NDT&E using shearography with impulsive thermal stressing and clustering phase extraction. *Opt Lasers Eng* 47(7):774–781
- Hung YY (1974) A speckle-shearing interferometer: a tool for measuring derivatives of surface displacements. *Opt Commun* 11(2):132–135
- Hung YY (1982) Shearography: a new optical method for strain measurement and nondestructive testing. *Opt Eng* 21(3):213–395
- Hung YY (1996) Shearography for non-destructive evaluation of composite structures. *Opt Lasers Eng* 24(2–3):161–182
- Hung YY (1999) Applications of digital shearography for testing of composite structures. *Compos Part B* 30(7):765–773
- Hung YY, Hovanesian JD (1982) Shearography-a new non-destructive testing method. *Amer Soc Non-Destructive Test* 40(3):A7–A8
- Hung YY, Hovanesian JD (1990) Fast detection of residual stresses in an industrial environment by thermoplastic-based shearography. In: 1990 SEM spring conference on experimental mechanics, Society for Experimental Mechanics, Albuquerque, USA, pp 769–775
- Hung YY, Shi D (1998) Technique for rapid inspection of hermetic seals of microelectronic packages using shearography. *Opt Eng* 37(5):1406–1410
- Hung YY, Wang JQ (1996) Dual-beam phase shift shearography for measurement of in-plane strains. *Opt Lasers Eng* 24(5–6):403–413
- Hung MYY, Long KW, Wang JQ (1997) Measurement of residual stress by phase shift shearography. *Opt Lasers Eng* 27(1):61–73
- Hung MY, Shang HM, Yang L (2003) Unified approach for holography and shearography in surface deformation measurement and nondestructive testing. *Opt Eng* 42(5):1197–1207
- Ibrahim JS, Petzing JN, Tyrer JR (2004) Deformation analysis of aircraft wheels using a speckle shearing interferometer. *Proc Inst Mech Eng G* 218(4):287–295
- Jackson ME (2004) Research and development of portable self contained digital phase shifting shearography apparatus to measure material properties through optical methods. Doctoral dissertation, Oakland University
- Kalms M, Osten W (2003) Mobile shearography system for the inspection of aircraft and automotive components. *Opt Eng* 42(5):1188–1196
- Kästle R, Hack E, Sennhauser U (1999) Multiwavelength shearography for quantitative measurements of two-dimensional strain distributions. *Appl Opt* 38(1):96–100
- Krivtsov VV, Tananko DE, Davis TP (2002) Regression approach to tire reliability analysis. *Reliab Eng Syst Saf* 78(3):267–273
- Lassahn GD, Lassahn JK, Taylor P, Deason VA (1994) Multiphase fringe analysis with unknown phase shifts. *Opt Eng* 33(6):2039–2045
- Lee JR, Molimard J, Vautrin A, Surrel Y (2004) Digital phase-shifting grating shearography for experimental analysis of fabric composites under tension. *Compos A: Appl Sci Manuf* 35(7):849–859
- Lee JR, Yoon DJ, Kim JS, Vautrin A (2008) Investigation of shear distance in Michelson interferometer-based shearography for mechanical characterization. *Meas Sci Technol* 19(11):115303
- Leendertz JA, Butters JN (1973) An image-shearing speckle-pattern interferometer for measuring bending moments. *J Phys E* 6(11):1107
- Li J, Xie X, Yang G, Zhang B, Siebert T, Yang L (2017a) Whole-field thickness strain measurement using multiple camera digital image correlation system. *Opt Lasers Eng* 90:19–25
- Li J, Dan X, Xu W, Wang Y, Yang G, Yang L (2017b) 3D digital image correlation using single color camera pseudo-stereo system. *Opt Laser Technol* 95:1–7

- LOBANOV LM, BYCHKOV SA, PIVTORAK VA, DERECHA VY, KUDERS' KYI VO, SAVYTS' KA OM, KYYANETS' IV (2009) On-line monitoring of the quality of elements of aircraft structures by the method of electron shearography. *Mater Sci* 45(3):366–371
- LØKBERG OJ (1987) Electronic speckle pattern interferometry. In: *Optical metrology*. Springer, The Netherlands, Dordrecht, pp 542–572
- MACY WW (1983) Two-dimensional fringe-pattern analysis. *Appl Opt* 22(23):3898–3901
- PEDRINI G, ZOU YL, TIZIANI HJ (1996) Quantitative evaluation of digital shearing interferogram using the spatial carrier method. *Pure Appl Opt* 5(3):313
- RŮŽEK R, BĚHAL J (2009) Certification programme of airframe primary structure composite part with environmental simulation. *Int J Fatigue* 31(6):1073–1080
- SHANG HM, HUNG YY, LUO WD, CHEN F (2000) Surface profiling using shearography. *Opt Eng* 39(1):23–32
- SIM CW, CHAU FS, TOH SL (1995) Vibration analysis and non-destructive testing with real-time shearography. *Opt Laser Technol* 27(1):45–49
- SIROHI RS (1984) Speckle shear interferometry. *Opt Laser Technol* 16(5):251–254
- STEINCHEN W, YANG L (2003) Digital shearography: theory and application of digital speckle pattern shearing interferometry, vol 93. SPIE press, Bellingham
- STEINCHEN W, YANG LX, SCHUTH M, KUPFER G (1994) Electronic shearography (ESPSI) for direct measurement of strains. In: *Optics for productivity in manufacturing*. International Society for Optics and Photonics, the International Society for Optical Engineering, Frankfurt, Germany, pp 210–221
- STEINCHEN W, YANG LX, KUPFER G (1996) Vibration analysis by digital shearography. In: *Second international conference on vibration measurements by laser techniques: advances and applications*, vol 2868. International Society for Optics and Photonics, the International Society for Optical Engineering, Ancona, Italy, pp 426–438
- STEINCHEN W, YANG LX, KUPFER G, MÄCKEL P, VÖSSING F (1998a) Strain analysis by means of digital shearography: potential, limitations and demonstration. *J Strain Anal Eng Des* 33(2):171–182
- STEINCHEN W, YANG L, KUPFER G, MÄCKEL P (1998b) Non-destructive testing of aerospace composite materials using digital shearography. *Proc Inst Mech Eng G* 212(1):21–30
- TAILLADE F, QUIERTANT M, BENZARTI K, AUBAGNAC C (2011) Shearography and pulsed stimulated infrared thermography applied to a nondestructive evaluation of FRP strengthening systems bonded on concrete structures. *Constr Build Mater* 25(2):568–574
- TARGOWSKI P, ROUBA B, WOJTKOWSKI M, KOWALCZYK A (2004) The application of optical coherence tomography to non-destructive examination of museum objects. *Stud Conserv* 49(2):107–114
- TESFAMARIAM S, GODA K (eds) (2013) *Handbook of seismic risk analysis and management of civil infrastructure systems*. Woodhead publishing, Elsevier, Cambridge, UK
- TOH SL, CHAU FS, SHIM VPW, TAY CJ, SHANG HM (1990) Application of shearography in nondestructive testing of composite plates. *J Mater Process Technol* 23(3):267–275
- TOH SL, TAY CJ, SHANG HM, LIN QY (1995) Time-average shearography in vibration analysis. *Opt Laser Technol* 27(1):51–55
- VEST CM (1979) *Holographic interferometry*. Wiley, New York. 476 p
- WANG Y, GAO X, XIE X, WU S, LIU Y, YANG L (2016) Simultaneous dual directional strain measurement using spatial phase-shift digital shearography. *Opt Lasers Eng* 87:197–203
- WU S, ZHU L, FENG Q, YANG L (2012) Digital shearography with in situ phase shift calibration. *Opt Lasers Eng* 50(9):1260–1266
- XIE X, XU N, SUN J, WANG Y, YANG L (2013a) Simultaneous measurement of deformation and the first derivative with spatial phase-shift digital shearography. *Opt Commun* 286:277–281
- XIE X, YANG L, XU N, CHEN X (2013b) Michelson interferometer based spatial phase shift shearography. *Appl Opt* 52(17):4063–4071
- XIE X, CHEN X, LI J, WANG Y, YANG L (2015) Measurement of in-plane strain with dual beam spatial phase-shift digital shearography. *Meas Sci Technol* 26(11):115202
- YAMAGUCHI I (2006) Phase-shifting digital holography. In: *Digital holography and three-dimensional display*. Springer, Boston, pp 145–171

- Yang L (1998) Grundlagen und Anwendungen der Phasenschiebe-Shearografie zur Zerstörungsfreien Werkstoffprüfung, Dehnungsmessung und Schwingungsanalyse. VDI Verlag, Düsseldorf
- Yang LX, Hung YY (2004) Digital shearography for nondestructive evaluation and application in automotive and aerospace industries. In: Proceedings of the 16 th WCNDT, Montreal
- Yang LX, Siebert T (2008) Digital speckle interferometry in engineering. In: New directions in holography and speckle. American Scientific, Stevenson Ranch, pp 405–440
- Yang LX, Steinchen W, Schuth M, Kupfer G (1995) Precision measurement and nondestructive testing by means of digital phase shifting speckle pattern and speckle pattern shearing interferometry. Measurement 16(3):149–160
- Yang L, Steinchen W, Kupfer G, Mäckel P, Vössing F (1998) Vibration analysis by means of digital shearography. Opt Lasers Eng 30(2):199–212
- Yang L, Chen F, Steinchen W, Hung MY (2004) Digital shearography for nondestructive testing: potentials, limitations, and applications. J Hologr Speckle 1(2):69–79
- Zhang Y, Li T, Li Q (2013) Defect detection for tire laser shearography image using curvelet transform based edge detector. Opt Laser Technol 47:64–71



# White Light Interferometry

14

Joanna Schmit and Anna Pakuła

## Contents

Introduction .....	422
Basics of Interference Fringe Formation .....	423
Coherence Influence on Two-Beam Interference .....	423
White Light Interference Fringes .....	430
Benefits Derived from White Light Interference .....	430
Fringe Analysis in WLI .....	431
White Light Microscope Interference System .....	436
Interference Microscope Objectives .....	437
Objectives Setup for Finding Fringes .....	440
Summary of Commercial Michelson and Mirau Interferometric Objectives .....	440
Polarization-Based WLI Microscope Systems .....	442
White Light Interference Systems Adapted to Samples Geometry .....	443
Large Complex Samples .....	443
Large Flat Samples .....	443
Large Cylinders .....	444
Objects Under Protective Glass or in Liquid Media .....	445
Matrix of Small Objects .....	446
WLI Fringe Analysis to Measure Special Features of Objects .....	447
Thick Film and Transparent Plates .....	447
Thin Film .....	448
Harmonically Moving Object .....	449
Surface Features Challenging to Measure .....	449
Object Imaging .....	451
Common Applications of WLI Systems .....	451
WLI as Auxiliary Signal to Measurement with Monochromatic Illumination .....	452
Roughness Measurement .....	452

---

J. Schmit

4D Technology, Tucson, AZ, USA

e-mail: [joanna.schmit@4dtechnology.com](mailto:joanna.schmit@4dtechnology.com)

A. Pakuła (✉)

Warsaw University of Technology, Warsaw, Poland

e-mail: [A.Pakula@mchtr.pw.edu.pl](mailto:A.Pakula@mchtr.pw.edu.pl)

---

Shape and Waviness Measurement .....	455
Examples of Applications .....	455
Other Systems Based on WLI .....	456
Spectrally Resolved White Light Interferometry .....	456
WLI in Literature .....	458
References .....	459

---

### Abstract

White light interferometry is a well-known measurement technique commonly used in 3D shape and roughness characterization of engineered and biological objects. In this chapter, the theoretical background of two beam interference is first given. This section highlights the basics of interference, its dependence on temporal and spatial coherence of the source, and how coherence is used in different types of measurement. Next, the basics of fringe formation is discussed, followed by a description of the most common white light interference analysis methods. Frequently used and commercially available white light interferometry systems based on microscopes with interference objectives, namely Michelson, Mirau, and Linnik, are discussed, followed by a description of setups built specially to adapt to chosen sample geometries – as large complex samples, flats or cylinders, immersed objects, and matrixes of small samples. Next is a discussion of the analysis of white light fringes used to measure special features of objects, i.e., thick and thin films, harmonically moving objects, involving surface features that are challenging to measure (dissimilar materials, complicated geometry, etc.). A special case of white light interferometric system that is combined with a spectrometer is also briefly described. Finally, common applications of white light interferometry, those where the technique is used directly or those in which the white light signal is auxiliary, are briefly described.

---

## Introduction

White light interference is a well-known optical phenomenon used widely especially in the microcharacterization of engineering and biological objects. It is employed with various measurement techniques; the two most important are white light interferometry (WLI), which is often called low coherence scanning interferometry (LCSI) or optical profilometry, and optical coherence tomography (OCT).

WLI is used to characterize 3D shape and roughness, optical thickness, and layer thickness of various objects; it is used in diverse fields such as tribology and semiconductor manufacturing. OCT is used mostly to characterize the inner structure of biological tissues. OCT is discussed in ► Chap. 15, “Optical Coherence Tomography for NDE.”

Both techniques employ broad spectrum light sources, which means that the temporal coherence of the source is decreased allowing for formation of fringes only over a small depth around the tested surface. This formation of very localized fringes allows for extension of applicability of classical laser interferometry

commonly used in optical metrology for measurement of smooth surfaces. This chapter is devoted to white light interferometry and its applications.

After some theoretical considerations of interference fringe formation, which is influenced by not only the temporal but also the spatial coherence of the source, practical applications of white light fringes for surface measurement proceeded by fringe analysis and system description is provided.

---

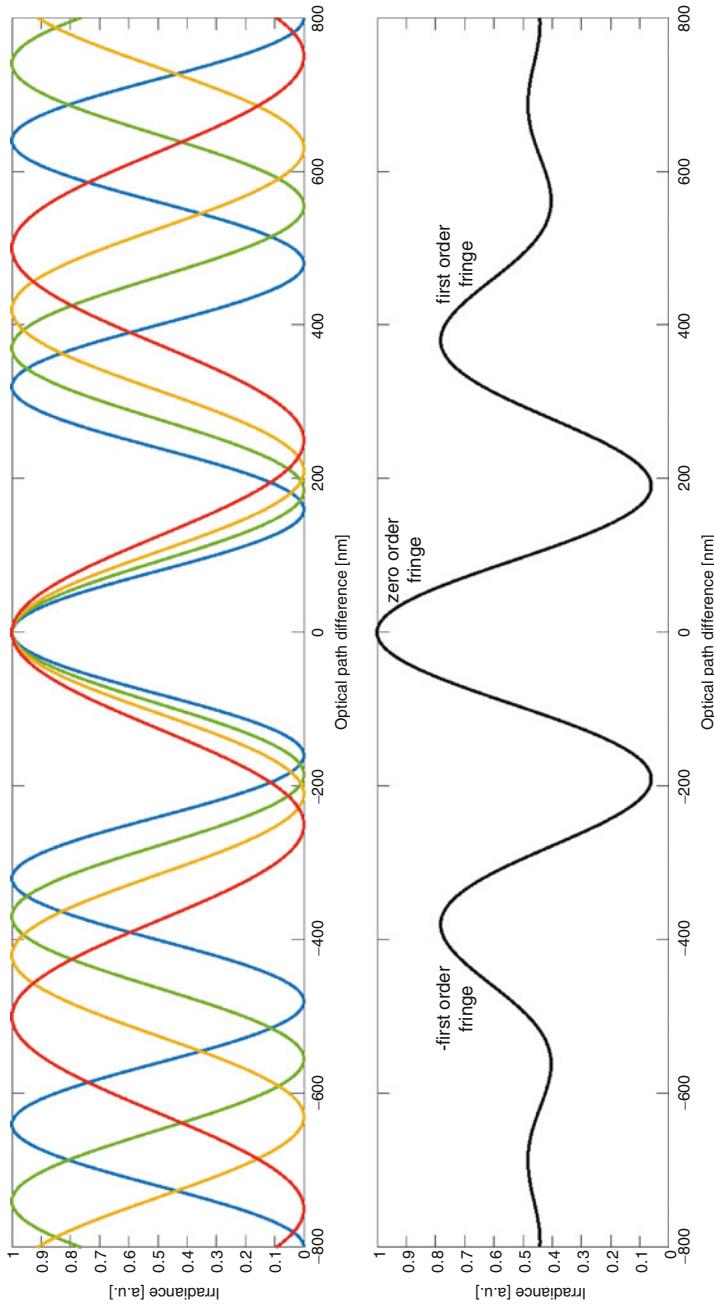
## Basics of Interference Fringe Formation

In this section, the theory behind the fringe formation from two-beam interference emitting from sources of different spatial and spectral extensions is presented for those interested in theory part. Later, a summary of the influence of spatial and temporal coherence of the light sources on the envelope of the fringes is given, followed by a review of literature describing methods utilizing different properties of sources for various applications.

## Coherence Influence on Two-Beam Interference

A key component of interferometric systems is the light source; typically, lasers (monochromatic and nearly single-point sources) have been used to generate interference patterns called fringes. The simplest perfect (single-point source and single wavelength) two-beam interference creates sinusoidal fringes within an almost infinite volume, which are referred to as nonlocalized fringes. However, interference also occurs when using broadband light sources, like white light-emitting diodes (LEDs); these are called low coherence sources. A broadband source refers to when the source spectrum emits more than a single frequency (or wavelength) of illumination causing the fall off of the contrast of interference fringes and limiting fringes to smaller volume. The fall off of the amplitude of the fringes is often referred to as a fringe envelope. These are called localized fringes. Figure 1 shows a comparison of fringe patterns for monochromatic and broadband light sources, which are also called high and low temporal coherence sources. The size of the emitting area of the source has a similar type of the effect on the fringes; analogously, single-point sources are called sources of high spatial coherence while sources of extended size have low spatial coherence.

To describe spatial and temporal coherence, the optical field propagating through the space needs to be first defined. A stationary optical field (called also wave) at position  $r$ , given in a function of time  $t$  is described by optical fluctuation  $E(r,t)$ . A detector placed at position  $r$  would measure the flux of the radiant energy of the statistically random field. This flux of the radiant energy is called irradiance,  $I(r) = \langle |E(r, t)|^2 \rangle$  (Palmer 1993). Irradiance is often commonly but incorrectly referred to as intensity. The position parameter  $r$ , because it is fixed, will be omitted in this chapter unless otherwise stated.



**Fig. 1** Interference fringes: (a) individual monochromatic fringes for a few single-frequency illuminations and (b) white light fringes localized in a smaller space volume

## Temporal Coherence

The main measure of temporal coherence, describing the extent of time for which the optical field fluctuates in unison at two instants of time, is a normalized *autocorrelation function*, also called *complex degree of temporal coherence* or *gamma coherence function* for which the equation is given in Eq. 1,

$$\gamma(\tau) = \frac{\langle E^*(t)E(t + \tau) \rangle}{\langle E^*(t)E(t) \rangle} \quad (1)$$

where  $E^*(t)$  is a field conjugate to  $E(t)$  and  $\tau$  is the time delay between field  $E^*(t)$  and its delayed replica  $E(t + \tau)$  (Saleh and Teich 2007).

The coherence function value carries information about the degree of correlation between  $E(t)$  and  $E(t + \tau)$ , whose fields in practice have to originate from the same source point; otherwise, there is zero correlation. The value of the complex degree of temporal coherence function cannot exceed unity, as shown in Eq. 2.

$$0 \leq |\gamma(\tau)| \leq 1 \quad (2)$$

The time delay for which  $|\gamma(\tau)|$  drops to  $1/e$  or  $1/2$  is referred to as *coherence time*  $\tau_c$ . Coherence time is the power-equivalent width of a coherence function, as shown in Eq. 3.

$$\tau_c = \int_{-\infty}^{\infty} |\gamma(\tau)|^2 d\tau \quad (3)$$

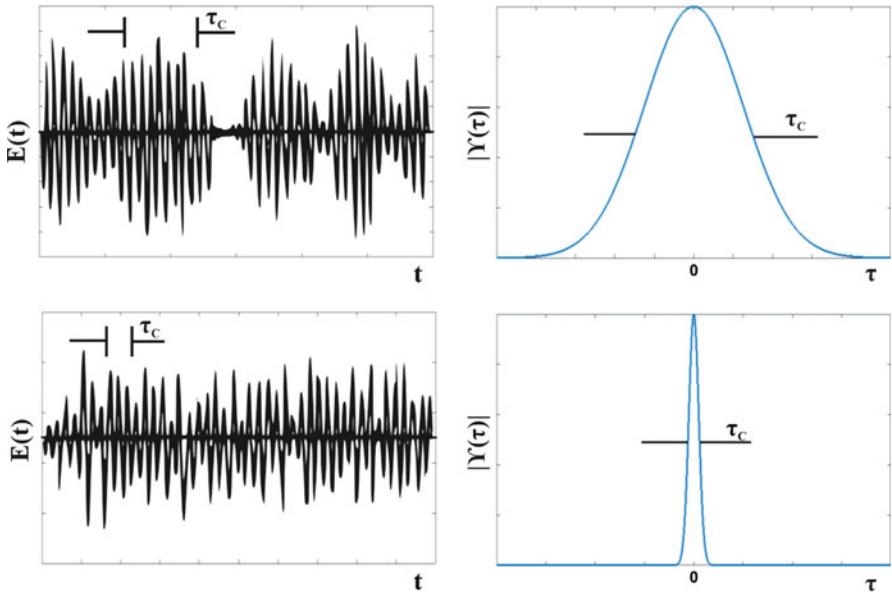
The dependence between the coherence function and coherence time is shown in Fig. 2.

Thus, the source of radiation can be characterized by the coherence time  $\tau_c$  or its equivalent coherence length  $l_c = c\tau_c$ , where  $c$  is the speed of light.

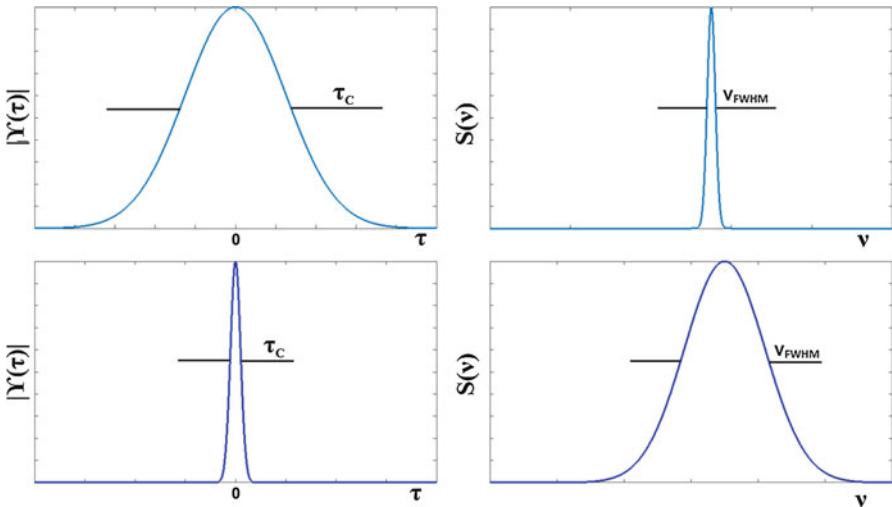
A monochromatic light source (described as the single angular frequency  $\omega_0$  and the single wavelength  $\lambda_0$ ,  $\omega_0 = 2\pi\nu_0$ , where frequency  $\nu_0 = c/\lambda_0$ ) can be characterized by a complex degree of temporal coherence equal to  $\gamma(\tau) = \exp(i\omega_0\tau)$  and  $|\gamma(\tau)| = 1$  for all  $\tau$ . If no other factors occur, then using such light will result in interference fringes of highest visibility regardless of time delay. On the other hand, if incoherent light is used, then only for  $\tau = 0$ , the complex degree can reach maximum and  $|\gamma(0)| = 1$ . Those two extreme problems are only theoretical and in practice the cases in between only should be taken into consideration.

The *gamma coherence function* is related to the source spectrum  $S$  by Fourier transform according to the Wiener–Khinchin theorem (Mandel and Wolf 1995), as shown in Eq. 4.

$$S(\omega) = \int_{-\infty}^{\infty} \gamma \exp(-i2\pi\omega\tau) d\tau. \quad (4)$$



**Fig. 2** The dependence between the coherence function and coherence time



**Fig. 3** The dependence between the spectral range and coherence time

The relation is shown in Fig. 3. The light of broad spectral range has a short coherence time, while a narrow band light source has longer coherence time. For monochromatic light, the spectrum consists of a single frequency  $S(\omega) = I\delta(\omega - \omega_0)$

**Table 1** Relations between coherence time  $\tau_C$  and spectrum profile (of spectral width given for Full Width Half Maximum  $\nu_{FWHM}$ ) for the most common cases.

Spectrum profile	Coherence time $\tau_C$
Rectangular	$\tau_C = 1/\nu_{FWHM}$
Lorentzian	$\tau_C = 1/(\pi\nu_{FWHM}) = 0.32/\nu_{FWHM}$
Gaussian	$\tau_C = (2\ln 2/\pi)^{1/2}/\nu_{FWHM} = 0.66/\nu_{FWHM}$

and  $\tau_C = \infty$ ,  $\Delta\omega = 0$ . Light with broader spectrum for which  $0 < \tau_c < \infty$  is considered as temporarily partially coherent.

The most common definition of *spectral width* or *linewidth*  $\Delta\omega_c$  of the source spectrum range is the full width at half maximum value (FWHM); however, generally spectral width can be derived from Eq. 5.

$$\Delta\omega_c = \frac{\left( \int_0^\infty S(\omega) d\omega \right)^2}{\int_0^\infty S^2(\omega) d\omega} \quad (5)$$

Equation 5 also takes into account the shape of the spectrum profile  $S(\omega)$  as they can be different for different sources. The relations between the *spectrum profile*, *coherence time*, and *coherence length*  $l_C = c\tau_C$  are shown in Table 1.

### Spatial Coherence

Spatial coherence is related to the lateral dimensions of the light source. As shown in the previous section, two beams emitted from a monochromatic point source will interfere at any location in space; however, there are no ideal point sources in the real world. The outcome of interference between two beams coming from an extended light source will suffer from spatial coherence effects which, unlike temporal coherence, do not depend on time delay between beams.

For the analysis of temporal and spatial coherence effects, optical field disturbances  $E(r_1, t)$  and  $E(r_2, t + \tau)$  originating in two different points in space are considered. The *complex degree of coherence*, which indicates the strength of the interference fringe modulation, is given in Eq. 6 as a cross-correlation coefficient that shows the degree of correlation between those disturbances,

$$\gamma(r_1, r_2, \tau) = \frac{\langle E^*(r_1, t) E(r_2, t + \tau) \rangle}{\sqrt{I(r_1) I(r_2)}}, \quad (6)$$

where  $I(r_1)$ ,  $I(r_2)$  are the irradiances  $I_{1,2} = \langle |E_{1,2}|^2 \rangle$  at points  $r_1$  and  $r_2$ , respectively. Finally, for  $\tau = 0$  and  $r_1 = r_2 = r$ , the irradiance equals complex degree of coherence at point  $r$ ;  $I(r) = \gamma(r, r, 0)$  (Saleh and Teich 2007).

The absolute value of the complex degree of coherence cannot exceed unity, as shown in Eq. 7.

$$0 \leq |\gamma(r_1, r_2, \tau)| \leq 1. \quad (7)$$

For the zero-time delay,  $\tau = 0$ , between the optical fields, the complex degree of coherence refers only to their position in space and is the measure of the degree of spatial coherence. If the degree of spatial coherence equals zero, the field disturbances at  $r_1$  and  $r_2$  are uncorrelated, while if the degree equals unity, then the fields are considered fully correlated. In practice, light coming from the point source is treated as spatially coherent; whereas, for extended light sources, its size determines the degree of the coherence.

Assuming that time delay  $\tau = 0$  and the path difference is much shorter than the coherence length  $l_C = c\tau_C$ , the light can be treated as monochromatic. In this case, a complex degree of coherence, as described by Eq. 8, is called normalized mutual intensity and describes the *spatial coherence* completely.

$$\gamma(r_1, r_2) = \frac{\langle E^*(r_1, t)E(r_2, {}^{2t}) \rangle}{\sqrt{I(r_1)I(r_2)}} \quad (8)$$

As in previous cases, Eqs. 2 and 7, the value of normalized mutual intensity is bounded between zero and unity. The value  $|\gamma(r_1, r_2)| = 1$  shows the complete correlation of the optical field in  $r_1$  and  $r_2$ , while  $|\gamma(r_1, r_2)| = 0$  states otherwise.

When discussing spatial coherence, it is not possible to omit the subject of coherence area, which is in the vicinity of  $r_1$  for which the function  $|\gamma(r_1, r_2)|$  is greater than 1/2 or 1/e. The coherence area is very important while characterizing random light, especially for describing its relation to the optical system. For example, if the area of coherence is greater than optical system's aperture ( $|\gamma(r_1, r_2)| \approx 1$  at all points), the light is considered coherent; otherwise, it is considered as spatially partially coherent.

### Interplay Between Temporal and Spatial Coherence Effects

Two partially coherent waves  $E_1$  and  $E_2$  (at the detector point  $r$  and time  $t$ ) of the irradiances  $I_1 = \langle |E_1|^2 \rangle$  and  $I_2 = \langle |E_2|^2 \rangle$  interfere, and the outcome of the interference is given by the following equation (Eq. 9)

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \operatorname{Re}\{\gamma_{12}\} \quad (9)$$

where  $\gamma_{12}$  is a normalized cross-correlation between the waves. Equation 9 may be rewritten as:

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} |\gamma_{12}| \cos \varphi \quad (10)$$

where the term  $\cos \varphi$  indicates the optical interference terms, where  $\varphi$  may be in range  $0-2\pi$ .

There are two very important boundaries of Eq. 10:

- $|\gamma_{12}| = 1$  represents two completely correlated (coherent) waves.
- $|\gamma_{12}| = 0$  represents two uncorrelated (incoherent) waves and there is no interference.

In general, both the above conditions can be met only in theory; however, with certain assumptions and omissions, they can be applied to practical situations. In general, the strength of the interference is measured by the contrast (or visibility) of the interference pattern as defined in Eq. 11

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad (11)$$

where  $I_{\max}$  and  $I_{\min}$  are the highest and lowest values that  $I$  takes with the change of  $\varphi$ . Up taking further transformations, it is possible to show that the contrast is proportional to the absolute value of normalized cross-correlation  $C \approx |\gamma_{12}|$ .

## Systems Based on Interplay Between Temporal and Spatial Coherence Effects

Although this chapter focuses on systems that use a broad spectrum, white light source that affects the fringe contrast via temporal coherence, some systems use a narrowband source with a high numerical objective or extended source that affects fringes mostly via spatial coherence (Abdulhalim 2001, 2006, 2012; Ryabukho et al. 2004; Lyakin and Ryabukho 2013; Cai et al. 2012; de Groot and de Lega 2007a). The controlled variation between spatial and temporal coherence in the system allows the system to adapt, for example, to measurements of small and large thicknesses of material layers (see sections “Thick Film and Transparent Plates” and “Thin Film”). Some systems are built on a controlled way of spatial rather than temporal coherence to avoid the influence of the inner aberrations of the system or to measure strongly curved surfaces (Küchel 2004; Gokhler and Rosen 2005a, b; Wang et al. 2002; Pavliček et al. 2009) or to move the coherence along the optical axis while keeping the object and reference mirror in one position (Rosen and Takeda 2000). It needs to be mentioned that not all systems based on low temporal or spatial coherence use white light illumination. The wavelength and the bandwidth of the source may vary, and generally any system based on low temporal or low spatial coherence is called, as accepted by International Standard Organization, the low coherence interferometer (LCI), and in the literature, the low coherence is often referred to as longitudinal coherence.

The next section describes the principles of white light interference systems based on temporal coherence and clarifies how this method can be used for surface topography measurement.

## White Light Interference Fringes

An optical interferometric system splits the wave into two wave components  $E_1$  and  $E_2$  using a beamsplitter, and a time delay is introduced between the different optical paths that the waves travel.

The irradiance of the resulting recombined field is measured as a function of the optical path difference (OPD). The contrast of the interference fringes created depends on the coherence properties of the source, as described in the previous section, and the OPD. For white light interference fringes in systems for surface topography measurement, fringes are localized around the zero optical path difference between both arms of interferometer as shown in Fig. 4. The zero optical path difference position with respect to the reference mirror corresponds to the best focus position of the sample. The fringe localization area is determined by the temporal coherence of the system, which is determined by the spectral bandwidth of the source. For white light interferometric systems, with commonly used LED source, this localization of fringes is within a few microns on both sides of the best focus position (Fig. 5).

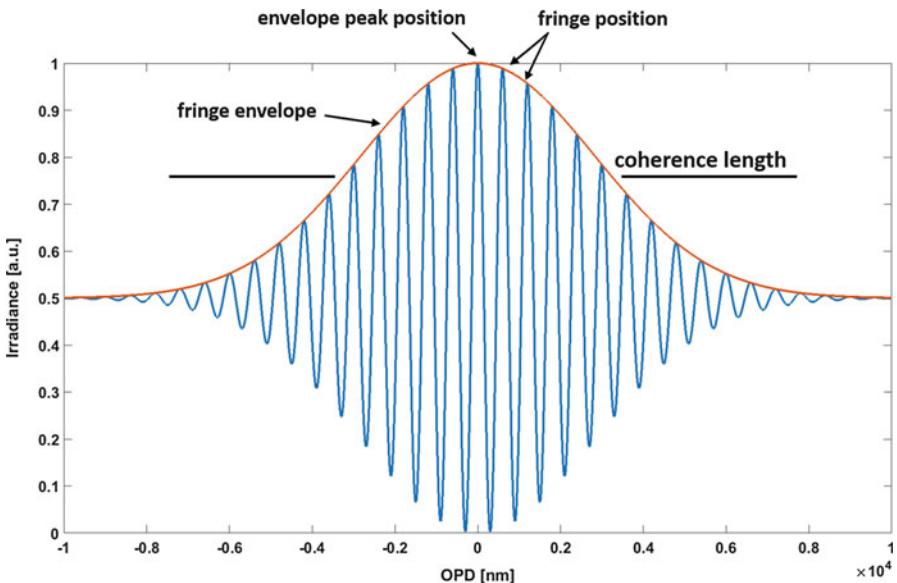
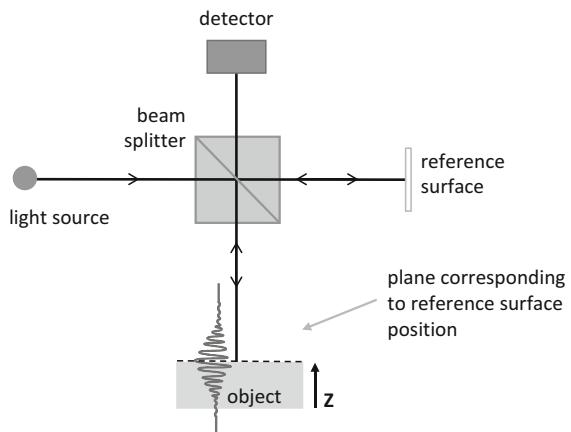
In order to further narrow the fringe envelope around best focus, systems with supercontinuum sources are build. Theoretically, the zero-order fringe, the fringe that indicates  $OPD = 0$ , is the highest amplitude irradiance fringe (or the lowest depending on the interferometer's configuration and the number of reflections in both arms of the interferometer) and the irradiance of other fringes decreases while the OPD increases. This decrease of fringe amplitude in white light interferometry, which is represented by the fringe envelope function as shown in Fig. 5, allows for the determination of the surface position where classical interferometry using laser or other highly coherent sources fails. Interference signal modeling is described extensively in the literature (de Groot and de Lega 2004).

## Benefits Derived from White Light Interference

The main drawback of using classical interferometry, which employs phase shifting interferometry (PSI) techniques and fringe analysis (Schreiber and Bruning 2007), is that it is limited in its ability to measure rough surfaces or surfaces with discontinuities. For surface discontinuities, the PSI measurement range is limited to  $\lambda/2$  or in practice to  $\lambda/4$ , where  $\lambda$  is the wavelength of the laser or other illumination source used. For samples exceeding this limitation (surface height difference between adjacent sampling points), fringe ambiguity occurs, and PSI fringe analysis is not able to assign proper height difference (Fig. 6).

By using a broadband spectrum light source (low temporal coherence) and axial scanning of the interferometer's head or sample, this limitation is eliminated. Simultaneous application of a broadband light source and axial scanning in a white light interferometer theoretically results in an unlimited measurement range. For each step of the axial scan, the irradiance at each pixel is recorded, and the distribution of the peak of the coherence envelope is found. The idea of measurement in WLI is shown in Fig. 7.

**Fig. 4** Simple interferometer drawing

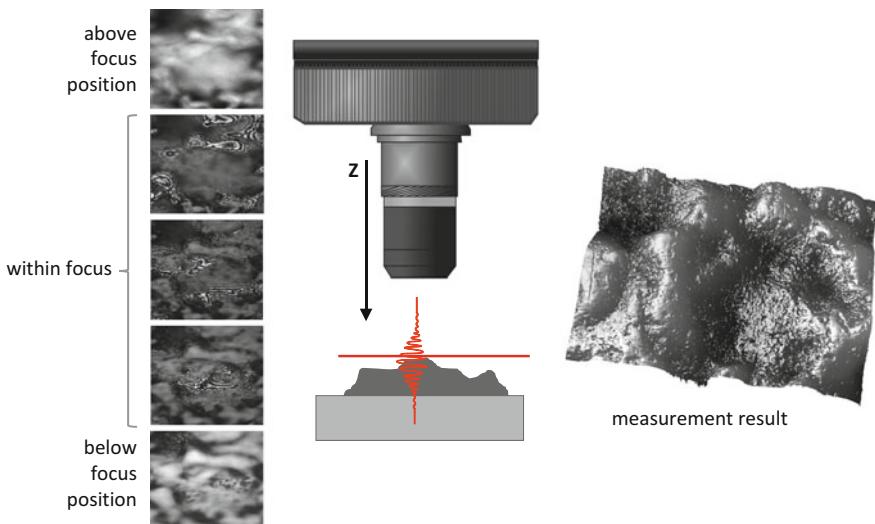
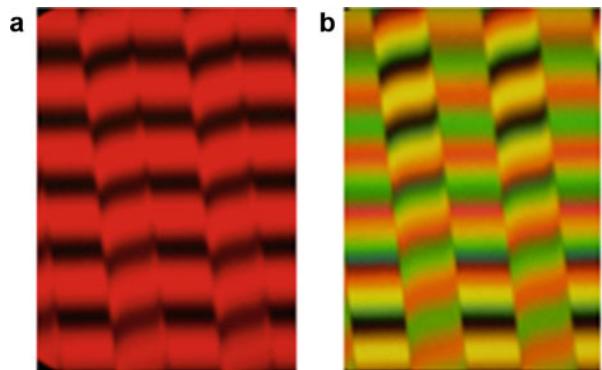


**Fig. 5** Localization of interference fringes within the coherence length of white light

### Fringe Analysis in WLI

As mentioned in the previous section, white light fringes can only be found within microns or tens of microns of the zero OPD, which indicates the sample position. Using this fringe localization requires axial scanning of either the sample or the whole interference objective through focus. During each step of a scanning motion, the irradiance at each pixel is registered. This irradiance is then analyzed in image postprocessing. Many methods exist for finding the best fringe localization position.

**Fig. 6** Fringe ambiguity issue for (a) classical laser interferometry and removed fringe ambiguity issue for (b) white light interferometry, for sample with step height exceeding measurement range for visible light illumination

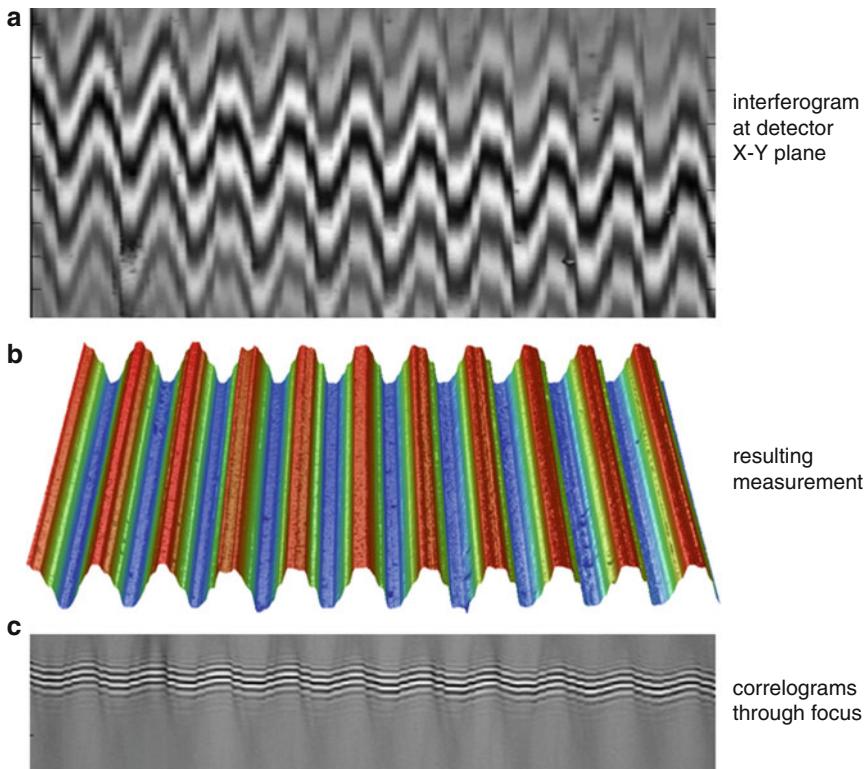


**Fig. 7** The idea of measurement in white light interference microscopy with axial scanning

Most of these methods are based on finding the highest point of the maximum fringe or the maximum of the calculated envelope of all fringes. Figure 8 shows an example of fringes as seen by the camera detector at the plane near focus and the resulting topography from the analysis of the interference fringes (correlograms) as observed during the scan through focus. The position of the localized fringes in the correlogram (even visually) clearly indicates the shape of the object. This section describes various methods of finding the position of fringes.

### Signal Processing of White Light Interferograms

In WLI, the object's shape can be obtained from the localization of the zero-order fringe registered during the axial scan at each point. Since 1980 (Balasubramanian 1982), many methods and algorithms have been developed that are based on the uniqueness of



**Fig. 8** Example of white light fringes near best focus as observed by the detector array, resulting measurement of structured surface and fringes (set of correlograms) as observed during the scan through focus by a horizontal line of detectors

WLI fringe distribution. Although the main characteristic for WLI is the highest irradiance recorded for zero-order fringe, it is very difficult to correctly calculate this exact position. Very often, the irradiance recorded for adjacent fringes is close to the zero-order fringe and the difference between them is lost due to relatively low noise level. The zero-order fringe position and amplitude can be affected by many factors, such as a dispersion imbalance between arms of the system, the phase change upon reflection either from the measured object's material properties or within the system itself, or even due to object tilt (Pfärtner and Schwider 2001).

These kinds of problems are the reason that in most cases it is more reliable to compute the envelope (modulation) of the fringes and determine its peak. Many ways are used to calculate the fringe envelope. The simplest way is to calculate the fringe envelope in the same way the modulation of fringes is determined in PSI. In this case, the OPD is changed during the scan motion along optical axis  $z$ , typically by a step that indicates a  $90^\circ$  change in beam phase between registered frames, but over a much longer scan range than with classical PSI. For final modulation determination, any PSI algorithm can be implemented.

Implementing a standard, 5-frame PSI algorithm (Hariharan et al. 1987; Schwider et al. 1983), the modulation would be calculated as shown in Eq. 12

$$\gamma(z) = \frac{\sqrt{(I_2 - I_4)^2 + (I_1 - 2I_3 + I_5)^2}}{\text{normalization}} = \frac{\sqrt{N^2 + D^2}}{\text{normalization}} \quad (12)$$

where  $N$  and  $D$  represent numerator and denominator of any PSI algorithm.

A useful review of PSI algorithms' properties that can be used for modulation calculation was done by Larkin (Larkin 1996a). It is also possible to use the Fourier transform technique to calculate the envelope with the method proposed by Kino and Chim (Kino and Chim 1990). This method involves computing the Fourier transform of the interference signal; in this technique, three lobes in the frequency domain are obtained. One of the sidelobes, positioned at the frequency of the fringes, is isolated and shifted to the center and the Fourier transform is applied again to calculate fringe envelope. The envelope's width is inversely proportional to the bandwidth of the source. Another method, proposed by Caber (1993), uses electronic hardware (amplitude demodulation in the hardware) to obtain the envelope of fringes.

Regardless of the method by which the envelope has been determined, the next step of WLI signal processing is determining the envelope's position. This can be done by employing different procedures, i.e., a curve-fitting procedure or envelope's center of mass calculation using Eq. 13:

$$h = \frac{\sum_{i=1}^{N-1} \gamma_{Zi} z_i}{\sum_{i=1}^{N-1} \gamma_{Zi}} \quad (13)$$

where  $\gamma$  represents the envelope function,  $z$  the axial position, and  $h$  the object's height.

The center of mass algorithm is relatively fast and computationally efficient, and for that reason it is often implemented in both interferometry and confocal systems for determining the maximum of the irradiance signal. However, to determine the maximum of the envelope function, like in WLI, this algorithm can only be applied for analysis of symmetrical signals. For the analysis of an asymmetrical signal, a piston is introduced for each point; however, this piston does not affect the whole measurement.

From all the methods used to determine the envelope's position, the most precise uses an achromatic phase shifter (Hariharan 1995, 1996). For each scanning position, the achromatic phase shifter shifts the fringes underneath the envelope. As a result, the constant modulation of the fringes is registered.

De Groot and Deck (de Groot and Deck 1995) showed that it is possible to find the position of the fringe envelope by processing the interference signal in the frequency domain. First, similar to Kino and Chim (Kino and Chim 1990), the Fourier transform is calculated and one sidelobe is isolated. The magnitude at each point of the sidelobe directly corresponds to the interference signal strength for the

given wavelength, while the phase represents the phase of the interference signal for this wavelength. They demonstrated that the phase slope of the spectral components of the white light interference can be used for finding the envelope's position with increased precision.

With recent increases in the performance of graphics processing units and the application of Compute Unified Device Architecture (CUDA), it is now possible to determine the position of maximum irradiance correlated to zero-order fringe instead of envelope determination. Proposed by Tomczewski (Tomczewski et al. 2013), least-squares polynomial curve fitting of the acquired data enables achieving high performance by parallel computation. This technique applies fitting of a cosine function to the central fringe only. This technique suppresses potential errors attributable to aberrations, diffraction effects, and unbalanced dispersion.

Application of alternative methods like wavelet transforms for WLI signal analysis is also possible (Recknagel and Notni 1998; Sandoz 1997; Yamada et al. 1994; Yatagai 1994). The main advantage of using this method is reduced noise; the disadvantage is longer processing time.

At first, white light interference microscopes were used to test smooth surfaces (Davidson et al. 1987; Kino and Chim 1990; Lee and Strand 1990). Gradually, applications were developed to measure rough surfaces (Häusler and Neumann 1993; Caber 1993), for which the speckle effect due to surface roughness had to be accounted for (Häusler and Herrmann 1993; Pavliček and Soubusta 2003) and possibly partially corrected (Wiesner et al. 2012).

While measuring the envelope position is the only method available to measure rough surfaces, measurement precision can be increased for smooth surfaces that may include discontinuities when the phase of the fringes under the envelope is determined in addition to the envelope peak position (Ai and Caber 1994; Harasaki et al. 2000; de Groot et al. 2002; Chen 2006). Methods based on combining position of envelope and fringes are often preferred over direct zero fringe order finding.

While this review of the literature describing algorithms for WLI is not complete, it does cover most of the commonly used analysis methods. Note that new methods are being developed all the time.

### **Sampling of White Light Interference Signal**

Measurement time in WLI strongly depends on axial scan rate and range. To reduce this measurement time, the sampling rate is often reduced, so that only a few axial points along the fringe envelope are detected (Creath 1997; Larkin 1996a, b). For example, applying PSI algorithms to calculate the envelope originally required four samples per fringe; however, if four samples per three or five fringes (odd number of fringes) are collected, faster but less precise data can be obtained. When scanning step size approaches the width of the fringe envelope, the sampled data becomes unusable. However, the width of the envelope can be widened by filtering the spectral bandwidth of the source and thus speed can be increased. Applying this method can increase measurement speed by as much as 23 times, which corresponds to 100  $\mu$  per second using a 60-frames-per-second camera frame rate (Schmit 2003). Even higher camera frame rates can be used to speed up the measurement further.

Issues of sampling for various WLI algorithms were widely discussed by a number of authors (de Groot and Deck 1993; Deck and de Groot 1994; Creath 1997; Hirabayashi et al. 2002; Larkin 1996a, b, Albertazzi and Dal Pont 2005).

While different algorithms allow for analysis of fringes of different surfaces, different interferometric systems allow for collection of fringes from different range of surfaces; these systems and their applications are described in sections “[White Light Interference Systems Adapted to Samples Geometry](#)” and “[WLI Fringe Analysis to Measure Special Features of Objects](#).<sup>”</sup>

### **Scanning and Vibration Error Self-Calibration**

A WLI measurement scan can be over 10 mm long and last more than a minute. Vibrations and scanner nonlinearity during such long scans may influence sampling during scanning and thus the measurement. Vibrations affect the interference signal itself, which adds additional error. Different methods have been developed to monitor the motion of the scanner and even vibrations independently from the measurement signal (Kang et al. 1999; Schmit and Olszak 2002; Olszak and Schmit 2003; Li et al. 2017; Mun et al. 2015; Tereschenko et al. 2016). Having precise information about the scanner motion helps avoid or at least reduce the resulting errors caused by unequal sampling. These methods improve both system accuracy and short- and long-term repeatability.

A very precise phase step can be calculated using temporal polarization phase shifting (geometric phase shifting) at each scanning step (Hariharan 1996). Alternatively, accurate phase steps can be calculated from spatial polarization phase shifting within each four pixels of a single frame captured with a polarization pixelated camera (Wiersma and Wyant 2013). More on this method can be found in section “[Summary of Commercial Michelson and Mirau Interferometric Objectives](#)” on polarization-based WLI microscope systems.

---

## **White Light Microscope Interference System**

One of the most common applications of white light interferometry is the measurement of object roughness and small structures on the object. Since there is a need to measure small areas at very high lateral resolution (de Groot 2016), microscope system is a natural choice. Ordinary optical microscopes are widely used in biological, medical, and material science to magnify images of small objects; their very fine lateral resolution allows for imaging of tiny structures. By modifying the objective and inserting an interferometer into the objective, an ordinary microscope can be converted to an interference microscope and surface-roughness and features of engineered microstructures can be measured with high precision using WLI principles. These 3D WLI microscopes measure surface topography over rectangular areas ranging from a few millimeters to down to the tens of microns in size with heights from a few nanometers to a few millimeters.

The optical setup of the interference microscope is similar to a common brightfield, reflected-light microscope where infinity corrected optics, commonly

used since the 1980s, and Kohler illumination, introduced in the 1890s, are typically used. Kohler illumination provides uniform illumination, good image contrast, and high spatial resolution by imaging the plane of the source at or near the aperture stop of the objective – unlike in previously used critical illumination where image of the source is imaged at the object.

The interference requires the beam from the light source to be split, and in an interference microscope, the beam splitting is most commonly done just under or just before entering the objective. The interferometric objective setups that are typically used are based on the Michelson, Mirau, or Linnik interferometers. This section describes these three objectives while section “[White Light Interference Systems Adapted to Samples Geometry](#)” provides information on customized objectives and systems for specific applications that expand the applications of white light interferometry for shape measurement.

## Interference Microscope Objectives

Interferometric objectives are basically variations of the Michelson interferometer. They are equal-path interferometers and if carefully balanced for dispersion and chromatic aberration can be used with even a white light illumination source. A common characteristic of these objectives is that the object and the reference mirrors are placed at the best focus position allowing for the maximum contrast of fringes, marking with interference fringes each point on the object when brought to focus.

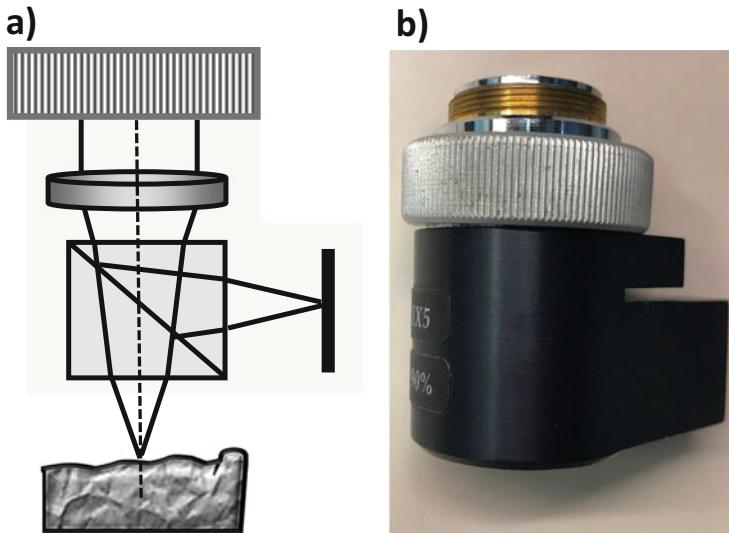
### Michelson Objective

A Michelson interferometer is comprised of a beamsplitter, placed under the objective, directing the beam to both the test object and the reference surface. A Michelson interferometric microscope objective must have a long working distance to fit the beamsplitter (typically a beamsplitter cube, but a tilted plate or pellicle is also used) between the objective and the surface. Because of this, Michelson interferometers are only used with low-magnification objectives having low numerical apertures and long working distances. The objective modified with a Michelson interferometer accommodates a reference mirror on the side of the objective and makes the objective, especially of lower magnification, bulky as shown in Fig. 9.

The constraints of the working distance of available brightfield objectives, being either too large or too small depending on magnification, led to innovative designs, with the biggest challenge being to develop interferometric objectives with very small and very large field of views.

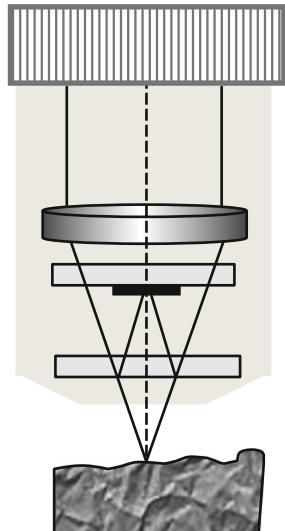
### Mirau Objective

The smaller working distance of the objective with magnifications above  $10\times$  does not leave enough space to accommodate the beamsplitter cube; Mirau interferometric setups (Mirau 1949), as shown in Fig. 10, use two identical thin glass plates under the objective to overcome this limitation.



**Fig. 9** Michelson objective

**Fig. 10** Mirau objective



One plate is coated on one side to act as a beamsplitter returning the beam toward the objective to bring to focus the beam on the other plate containing a small reflective spot that acts as the reference surface. The plate with the reference spot acts also as a compensating plate.

Over this past decade, commercial Mirau objectives have reached even  $100\times$  magnifications. Mirau setups are not very useful at magnifications of less than about  $10\times$  because at these lower magnifications the reference spot obscures too much of

the aperture. The reference spot has to be larger than the field of view of the objective since it is a surface conjugate to the best focus plane of the object.

A new, low-magnification interference objective, based on a variation of the Mirau setup, has recently been developed (de Groot and Biegen 2015). It allows for topography measurements over a field of view as large as 20 mm x 20 mm. This low-magnification objective replaces the reflective reference spot with a partially transparent reference plate; this plate, however, needs to be slightly tilted in order to avoid unwanted light being reflected back to the system (see section “[Large Flat Samples](#)”). This objective is more compact than a Michelson setup and still maintains fairly good fringe contrast.

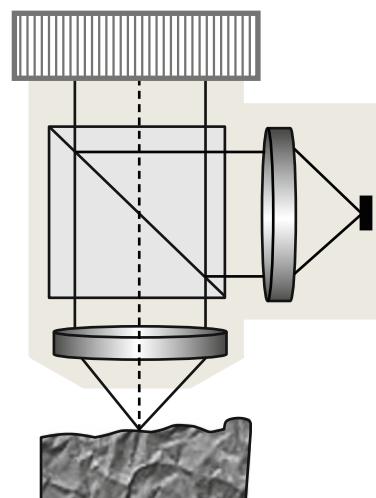
Sometime, for the most precise measurements, interferometric objectives are designed to work with polarized light (see section “[Summary of Commercial Michelson and Mirau Interferometric Objectives](#)”) but the polarized Mirau objective (Schmit and Hariharan 2007) is most challenging to produce.

### Linnik Objective

The Linnik setup (Linnik 1933) (Fig. 11) allows an interferometric objective to be set up for any magnification from two identical brightfield objectives.

This setup is often found in laboratories. The Linnik objective used to be the only way to make a high-magnification interferometric objective before the 100× Mirau objective was available on the market. Michelson and Mirau interferometric objectives have both a shorter working distance and a lower numerical aperture than a Linnik setup, which uses regular brightfield objectives. This system is often useful in polarization-based systems that employ orthogonal polarization between the arms of the interferometer. However, Linnik interferometers are more difficult to align as their setup is comprised of a beamsplitter, two matched microscope objectives and a reference mirror in order to provide a wavefront with minimum aberration and

**Fig. 11** Linnik objective



maximum fringe contrast. The entire reference arm, including the objective and distance before and after the objective, provides path-length matching in order to obtain white-light, low-coherence, fringes. The coherence effects in the Linnik microscope objective are described in the literature (Abdulhalim 2001) where researchers modified the objective to obtain different system benefits, such as high lateral resolution (Niehues et al. 2012) or long working distance (Hu et al. 2009).

To minimize the influence of dispersion on fringe contrast in Michelson and Linnik interference objectives, special beamsplitter cubes for white light interferometry (Farr and George 1992; Koudelka and Burge 2004; Hlubina et al. 2008) can be designed and fabricated.

## Objectives Setup for Finding Fringes

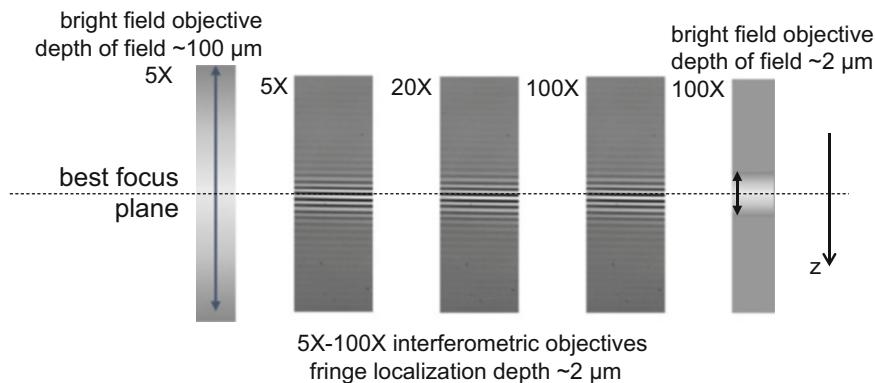
In order to properly setup the interferometric objective, not only does the objective need to be focused on the object, but also fringes have to be formed when the object is in focus. To obtain this double focus, the object first needs to be placed in focus and then the reference mirror needs to be moved to the best focus position. Only then is the zero optical path difference (OPD) between the reference mirror and the object position in the arms of the interferometer achieved and fringes of the best contrast created. Reference mirror position is typically set by the factory, so that the user when focusing on the object will also see fringes. Thus, fringes are focus indicators, and algorithms are developed to find the position of fringes for each point on the object during the scan through focus. In this way, precise information about object shape is obtained.

## Summary of Commercial Michelson and Mirau Interferometric Objectives

Table 2 summarizes typical parameters of commercial interferometric objectives and serves as a guide for matching the objective to the object measurement needs. When choosing the objective, the user has to understand what the required optical

**Table 2** Common characteristics of interference objectives (note that field of view and lateral sampling depends on detector's array and pixel size and will be different if field of view magnification lens is other than 1×)

Objective type	Magnification	Numerical aperture of the objective	Optical resolution [μm]	Field of view [mm × mm]
Michelson	2.5×	0.075	4.5	4.4×4.4
	5×	0.13	2.6	2.2×2.2
Mirau	10×	0.3	1.18	1.1×1.1
	20×	0.4	0.88	0.55×0.55
	50×	0.55	0.64	0.22×0.22
	100×	0.8	0.45	0.11×0.11



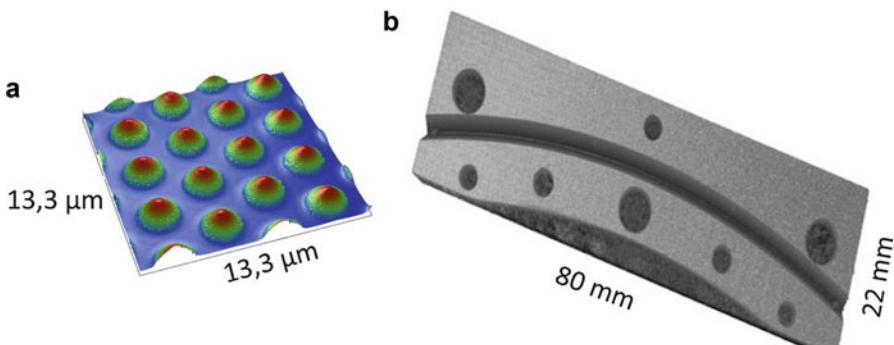
**Fig. 12** White light interference fringe localization for different magnification objectives and relative depth of field of objectives

resolution/lateral sampling would be as well as the field of view of the object's measurement. Sometimes it is difficult to get both in a single measurement.

One of the important parameters is magnification, which is determined as the ratio of the object size to its image size at the detector plane (not the display screen). The magnification can be modified by a choice of the objective and also a field lens typically from  $0.5\times$  to  $2\times$  magnification. Numerical aperture of the objective (NA) determines the optical resolution. Numerical aperture, and thus optical resolution, typically increases with the magnification of the objective. However, in digital systems, the resolving capability depends also on lateral sampling of the object. The measured area of the object (field of view) and lateral sampling thus depend on system magnification and the detector. One parameter that stays constant for white light illumination is the vertical resolution of each objective. Because of the broadband source, the temporal coherence dominates and the width of fringe envelope for each objective is about a few microns regardless of objective's depth of field. In Fig. 12, the white light interference fringes for different magnification objectives and their relative depth of field are shown.

Overall shape measurement can be measured with  $2.5\times$  and  $5\times$  objective. Roughness measurement requires a minimum  $10\times$  magnification objective. Very fine roughness, tiniest (micron size) structures and highest slopes, may require  $100\times$  magnification objectives. Examples of measurement with different magnification objectives are shown in Fig. 13. The maximum measurable object height is determined by the scanning range of the system and typically is between 1 and 10 mm. Objectives from Table 2 are parafocal with each other (45 mm from the objective's shoulder to the focus plane) and can be placed on a microscope's turret, which allows for a wide range of measurements in one system.

Other, less common, interferometric objectives and systems are available or are described in the literature and serve to solve specific measurement issues such as sample size, shape, or simply sample accessibility. These objectives have typically longer parafocal length and have to be mounted separately. Some of these



**Fig. 13** Examples of surface topography of (a) patterned sapphire structure in LED-making process measured with 115 $\times$  and (b) metal surface as measured with 2.5 $\times$  objective in a few overlapping measurements

customized solutions are described in section “[White Light Interference Systems Adapted to Samples Geometry](#)”. The main principle of white light measurement in these systems stays the same but the interference objective or system varies from those already described.

### Polarization-Based WLI Microscope Systems

Some WLI microscope systems use polarization to benefit measurement. Using polarization in interferometry is not new, as (Hariharan 1996) proposed a WLI system where the light is orthogonally polarized in both arms of interferometer and two types of phase shifting – mechanical and achromatic (also called geometric) – are used. Mechanical phase shifting is used to scan the object through focus; however, at each scan position, achromatic phase shifting is introduced to move fringes under the envelope. Achromatic phase shifting is then used to calculate precisely the modulation at each scan position since modulation is the same for all phase-shifted frames at a given mechanical step. Simultaneously, the fringe phase position is calculated from the achromatic phase shifting, which gives information about the scan step and allows for the scan steps to be larger than 90°. In addition, the information about the steps allows for measurement correction due to scan variations. The drawback to this setup is that collecting a few frames at each mechanical scan is quite time-consuming.

To solve this, Wiersma and Wyant proposed using a pixelated camera (Wiersma and Wyant 2013) that would allow for instantaneous calculation of phase at each point from four neighboring pixels in one frame instead of multiple frames. Thus, the system with polarized interference objectives and pixelated camera can allow for fast and vibration-insensitive white light interferometry measurements with a very small decrease in lateral resolution.

Polarization WLI microscope systems with a pixelated camera commonly use a Linnik objective as it is the easiest to modify the system to obtain two orthogonally

polarized beams. Some notable exceptions include (Schmit and Hariharan 2007), who suggested a design for a Mirau polarization objective with wave plates with a few follow-up designs (Lyulko et al. 2013; Bouchal et al. 2015; Tapisouw et al. 2013). Although technologically challenging, a few Mirau polarization objectives have been built (Wang and Liang 2016) and placed in systems with a pixelated camera for fast and vibration-insensitive measurements of stepped surfaces.

---

## White Light Interference Systems Adapted to Samples Geometry

WLI microscope systems are commonly built to measure roughness, structure, and shape of small objects. However, some sample's geometry requires system modification or an entirely new system. This section gives a few examples of systems that provide solutions for measurement of large samples, samples under protective glass, or for parallel measurement of matrix of samples.

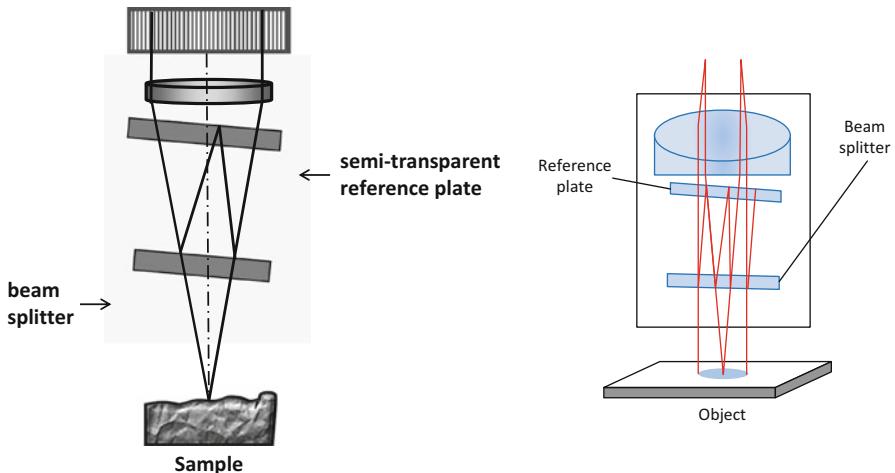
### Large Complex Samples

Typical WLI microscope systems are not able to accommodate large samples. Testing samples like orthopedic medical implants, rotors and camshafts, or aircraft engine turbines require a special mechanical design to allow for placing large pieces under the objective. In addition, samples can be placed on rotational stages and the whole microscope head can be tilted  $\pm 45^\circ$  around the sample allowing for the measurement of beveled edges and angled surface as well as many places otherwise not accessible with a typical setup. Special endoscope-type objectives can be built (Lindner 2002) in order to look deep into the small drilling of the object. Other systems can be made small and be brought to and placed on the large sample like a print roller or large telescope mirror, but vibration may be an issue for the measurement, unless a polarization WLI with pixelated camera is used.

### Large Flat Samples

Measuring large areas of an object can be achieved in a few different ways. Measurements of adjacent areas can be put together into one measurement by stitching slightly overlapped measurements. Alternately, a special wide-field objective can be used (de Groot and Biegen 2015). The stitching software is available on many systems and can be fully automated or the user can manually shift the object on the stage for consecutive measurements. Stitching works best on surfaces with easily identifiable structures.

The wide-field interference objective is made of two partially transparent plates placed under the objective like in the Mirau design but with two differences. First, there is no small reflective spot acting as a reference mirror; rather, the whole surface of the upper plate partially reflects the beam, creating the reference beam. Second, the



**Fig. 14** Wide-field interference objective

plates are partially tilted so as to avoid reflections from them going back to the objective. Similar tilts and equal path design are used in large aperture interferometers in order to avoid the presence of unnecessary bouncing beams that reduce the fringe contrast. The scheme of wide-field interference objective is shown in Fig. 14.

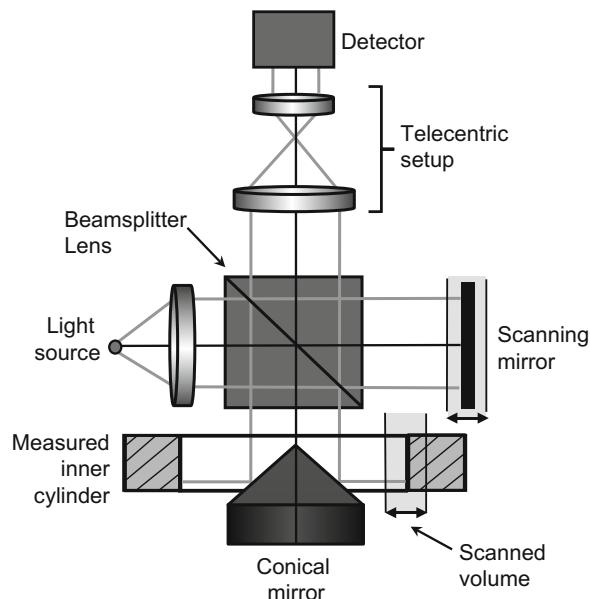
This type of design, rather than Michelson or Mirau, allows for a larger working distance still with a compact design, and it avoids the large central obscuration from a Mirau objective. Low-magnification objectives ( $2\times$  and below) of longer para-focal lengths (60–300 mm) can be modified in this way, and the field of view can reach a 6–17 mm view with a  $1\times$  tube lens. Optical resolution and sampling is on the order 7–30  $\mu\text{m}$ . Because this objective has a very low numerical aperture and coarse sampling, it is not suitable for measuring roughness and high slopes. It is mostly used for the measurement of flat mechanical surfaces.

Some WLI systems are not based on a typical microscope but are specially designed to use high aperture, long working distance telecentric objectives that allow for the measurement of flatness, thickness, and parallelism over large fields of view even up to 30 mm (Bauer 2003). These systems can be used for the measurement of mechanical components like gauge blocks, distance plates, optical plates up to 0–100 mm tall.

## Large Cylinders

Michelson objectives with very long working distances can be modified to “look” at an object located not beneath but to the side of the objective by switching the place of the reference mirror with the object. Such objectives can look at the inner walls of engine cylinders or bores at a depth of 150 mm, and, combined with a theta rotational stage, enable automated inner-cylinder surface stitching. For smaller inner cylinders (but also

**Fig. 15** Measurement of inner cylinder using cone mirror



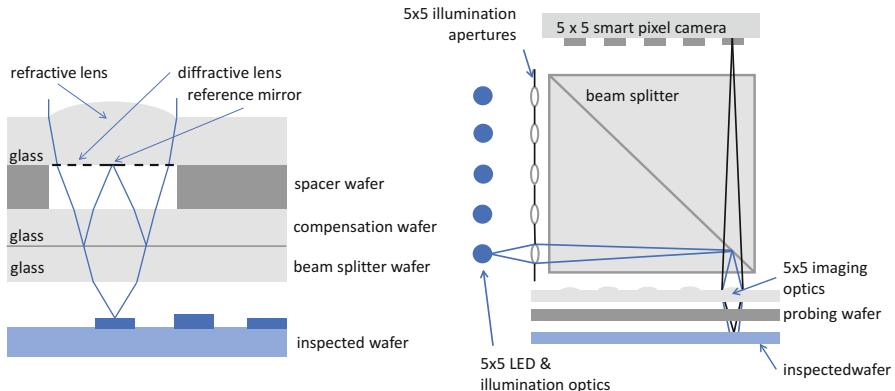
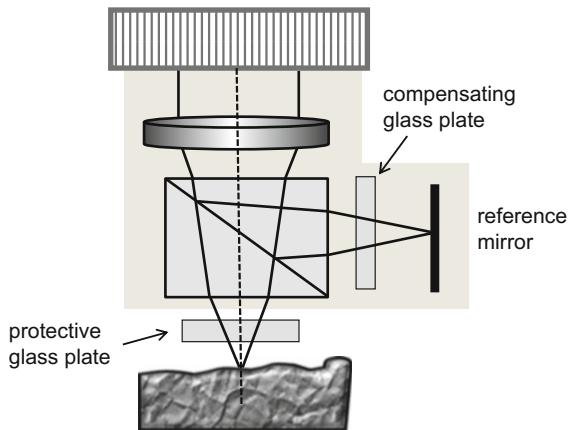
outer cylinders), a setup with a cone mirror inside or outside of the cylinder can be used (Albertazzi et al. 2008) as shown in Fig. 15. Some WLI systems can use an endoscopic probe for measuring the inside of small cylinders. WLI setups can also use a spherical instead of flat reference mirror to measure deviation of objects like a cone valve (de Groot and de Lega 2003). Other modifications of the interferometer for the measurement of more complicated shapes are possible (Lindner and Thominet 2005).

### Objects Under Protective Glass or in Liquid Media

Some objects like MEMS or MOEMS devices are protected by a glass cover or are immersed in liquid, like microfluidic devices. Testing these samples can be problematic because when they are placed under a regular interferometric objective, the fringes may have very weak contrast or be completely invisible due to the dispersion introduced by the cover glass or liquid dispersion. In these cases, a plate of equivalent optical thickness of suitable material must be inserted into the reference arm of the interferometer to balance optical path in both arms for fringes to regain good contrast (Han et al. 2005). Michelson and Linnik objectives are much easier to modify than the Mirau objective (see Fig. 16).

In order to increase further fringe contrast, or if the dispersion cannot be perfectly balanced, the illumination system can be separated from the imaging system. Low numerical illumination can be provided from the side of the beamsplitter underneath the objective, while the imaging optics brings the benefit of some numerical aperture of the objective.

**Fig. 16** Michelson WLI system with side illumination and compensating plate in the reference arm



**Fig. 17** Design of the wafer interferometer: Mirau configuration (left) and inspection setup concept (right)

## Matrix of Small Objects

Wafer-level manufacturing in the semiconductor industry also needs fast shape and deformation measurement of a large array (matrix) of etched patterns. An efficient way to measure these large arrays uses 25 Mirau-type interferometers, also built on a wafer (Kujawińska et al. 2010; Styk et al. 2010). To maintain the compact structure of each interferometer, refractive elements together with diffractive optical elements (DOE) are used. A  $5 \times 5$  matrix of smart-pixel cameras, which detects and demodulates the interference signal, performs optical lock-in detection of the interference signals at the pixel level (Fig. 17). The smart-pixel approach is applied for massive parallel electro-optical detection and data reduction. Other types of interferometer designs built on wafers are possible.

## WLI Fringe Analysis to Measure Special Features of Objects

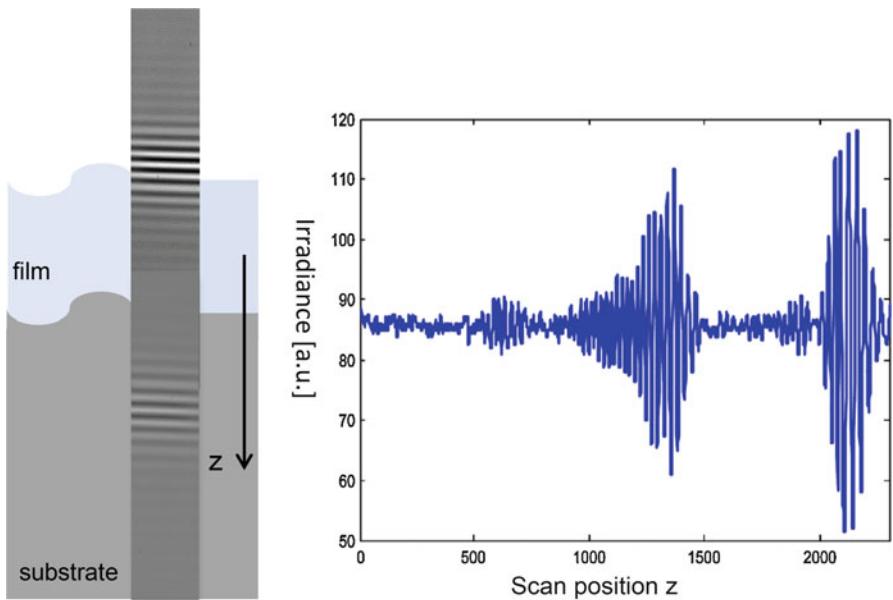
Besides surface topography measurement, which is the main application of WLI, systems can be used to measure transparent layers on samples. Since these transparent layers modify the fringe signal (multiple beam interference can be observed), a special algorithm needs to be used and often the temporal or spatial coherence of the source needs to be altered. Some features of the sample itself also can affect the fringe signal and thus the measurement. Thus, it is important to know how to recognize potential errors in the measurement and the ways to reduce them. This section helps with just that and in addition reveals that not only static samples can be measured but these in harmonic motion also.

### Thick Film and Transparent Plates

The narrow localization of fringes in white light interferometry allows for the measurement of not only surface topography but also the thicknesses of transparent layers roughly between 2 and 200  $\mu\text{m}$ . Measuring protective coatings on mechanical surfaces or a dissolvable transparent layer containing medicine on surfaces such as stents or implants is often important. For the measurement of transparent layers on the object, typically no special modification to the system hardware needs to be done; rather, software changes to analyze fringes affected by reflection from film interfaces is sufficient. Every interface of this coating, called a film, acts as reflective surface and the wavefront reflected from each interface will create interference fringes in the form of two separated sets of white light fringes. A simple technique for finding the relative position of the peaks of the fringe envelopes can be implemented to find the thickness of the film. The relative position of these two envelopes depends on the dispersion of the material. Their separation will depend not only on the index of refraction for a given wavelength  $n(\omega)$ , but also on the group index of refraction  $N(\omega) = n(\omega) + dn(\omega)/d(\omega)$  of the material, which is a measure of material linear dispersion. Figure 18 shows two almost clearly separated sets of fringes formed for the air/film and film/substrate interfaces. Measurement of coatings using white light interferometry provides the great benefit of not only measuring the thickness of the film but the whole topography of the top and bottom interface.

However, the measurable film thicknesses will depend on the spectrum of the source (temporal coherence), the dispersion of the film, and the numerical aperture NA of the objective. Dispersion effects will be stronger for sources with a wider spectrum and objectives with a higher NA. Thus, for thicker films, the bandwidth of the light should be reduced (even to 10 nm) and an objective of lower NA should be selected (Safrani and Abdulhalim 2011, 2012). White light interferometry has also been used to measure the thickness of oil film on water (Sun et al. 2005).

Dispersion introduced by a film may widen the envelope, the position of the fringes may shift under the envelope, fringes may lose contrast, or the period of the fringes may change with the  $z$  position. A good review of these effects can be found in (Pavliček and Soubusta 2004). Rough surfaces or interfaces of films may create a



**Fig. 18** Two almost clearly separated sets of fringes formed for the air/film and film/substrate interfaces

large speckle effect that distorts fringes and may not allow for reliable measurement, so caution needs to be taken when measuring films on rough surfaces.

Similar techniques are used for transparent plate thickness or their dispersion measurement (Haruna et al. 1998). Since the scan of a plate (thickness up to a few millimeters) takes a long time, a few techniques were developed which, by clever control of temporal or spatial coherence, create localized fringes at two or more planes (Gokhler and Rosen 2005b; Salsbury and Olszak 2017), which allows for faster plate thickness measurement.

## Thin Film

When the film is thinner than the coherence length of the light in the system, the three beams (from reference mirror and the two interfaces of the film) will interfere with each other and significantly (and in nonlinear fashion) affect the fringe envelope and the frequency of fringes (Hariharan and Roy 1996; Roy et al. 2005). Simply finding the envelope peak position is no longer valid. Fringes have to be analyzed differently, frequently involving signal modeling. Often the Fourier transform of the measured signal is used to calculate the so-called spectral phase (Kim and Kim 1999). The spectral phase for the thin-film interference is dependent not only on film thickness but also on the film's material property. To measure the film thickness, the

spectral phase for different film thicknesses is modeled and then compared to the measured phase in the process of best fitting measurement. Since this spectral phase is affected by the system's internal dispersion, the measured film's spectral phase needs to be first corrected for. Spectral phase of system dispersion is measured in the process of system calibration.

Many sophisticated algorithms were created for film thickness measurement (Mansfield 2006; de Groot and de Lega 2007b; de Lega and de Groot 2008a, b; Wan 2009; Kim et al. 2017; Yoshino et al. 2017; Fay and Dresel 2017). These algorithms are used for films of thickness from a few microns down to 100 nm. For films of optical thickness less than 100 nm, the sensitivity of the method drastically decreases (Dong 2012). And again, when using interferometry to measure thin or thick films, not only is thickness measured but also top and bottom profiles.

Sometimes not the film thickness but gaps between surfaces need to be measured; like in silicon wafer manufacturing, the gap between the mask and substrate needs to be known. To measure the gap, methods similar to thin and thick film measurements can be used but sometimes some modification in the system are done to reduce the influence of dispersion of the plates on the measurement of the gap (Xu et al. 2009).

## Harmonically Moving Object

WLI systems can also be modified to measure the motion of periodic oscillation of samples like micromechanical systems: MEMS devices or capacitive pressure sensors (Petitgrand et al. 2001; Bosseboeuf and Petitgrand 2003; de Groot 2006, 2008; Shavrin et al. 2013; Chen et al. 2009; Heikkinen et al. 2013, 2015; Kassamakov 2008). This modification requires a two-channel signal generator to drive the sample's harmonic motion and stroboscopic illumination. The sample is illuminated with short pulses at the frequency synchronized with the sample movement. The light pulse is short compared to the period of the sample motion and thus the sample appears stationary. By changing the relative phase between the signals driving the sample and the illumination, the sample's topography maps at different phases of its motion appear stationary and are measured. These systems can measure samples vibrating at frequencies from a few Hz up to a dozen MHz and with newer supercontinuum sources (Hanhijärvi et al. 2012, 2013) even up to a few tens of MHz. The stroboscopic pulse can be as short as 50 ns, but with special picosecond supercontinuum sources (Novotny et al. 2014) even much shorter pulses are possible. These super-short pulses potentially could be used to measure vibrations with frequencies on the order of hundreds of MHz.

## Surface Features Challenging to Measure

Some challenges still exist in using WLI, ranging from material composition to edge effects. Some of these effects may be minimized by proper choice of objective and careful system alignment.

## Dissimilar Materials

First of all, if a sample is made of different materials within the same field of view, each material, due to its different spectral phase change on reflection, will affect the fringes in its own way, and the relative heights between these areas will carry some offset (Schulz 1951; Park and Kim 2001). Metals have the strongest effect on the measurement and may introduce offset of a few tens of nanometers (Harasaki et al. 2001); however, this error may be ignored depending on sample's vertical range. Semiconductor materials will have a smaller effect while a dielectric will have no effect. Caution needs to be exercised if the sample has some kind of transparent coating; nonlinear effects were described in the previous section.

## Sharp Edges

One of the annoying effects in white light interferometry is overestimation of the height near some sharp discontinuities on the sample. This error source was named the “batwing” effect due to its characteristic profile obtained during measurement of structures with rectangular cross-section profiles. This effect (Harasaki and Wyant 2000) is present on structures when measured with light of coherence longer than object's height discontinuity. A similar effect may be observed when measuring a structure with a sinusoidal profile, although here the high slopes may not be resolved introducing large noise, but both effects may be reduced by choosing a large numerical objective – see section “[Roughness Measurement Issues](#).”

## V Grooves

Some structures on the object may be overestimated due to high local slopes reflecting the trapped light between them returning to the objective via an indirect path due to single reflection. This change of path creates an interference signal for apparent longer depth of the object (Gao et al. 2006, 2008). To check if the measurement is affected by multiple reflections, the sample can be tilted and measurements compared.

## Narrow Structures

Besides geometrical effects, diffraction effects can affect measured fringes. Diffraction effects are strong for narrow and sharp features on the object. A very good example of diffraction is ghost fringes; these fringes are created at the entrance of void areas in narrow trenches. These ghost fringes can be mistaken for the proper fringes at the bottom of the trench, and the strength of these ghost fringes can be even much greater than the strength of the trench bottom fringes. The presence of ghost fringes was noticed on an interference microscope and then also simulated for trench widths narrower than 10 µm (Schmit et al. 2003; Tavrov et al. 2005) and other small structures (de Groot et al. 2008). Such structures need to be analyzed with great caution and choice of the algorithm may be important.

## Object Imaging

Although in white light interferometry the fringes are present when the system is focused on the sample, however, during a scan not only can object shape be measured but also its image can be retrieved by analyzing fringes (de Groot 2005). The final image can be presented as an extended focus image (best imaging at each point brought into one plane).

There are some systems that also can image a sample in true color and not just in grayscale. These systems require color illumination or a color camera – with red, green, blue illumination or detection (Machleidt et al. 2012; Beverage et al. 2014; Schmit et al. 2015). Color imaging often requires good side illumination, which can be provided by using a ring illuminator (Schmit 2015).

A color camera also can be used for surface topography measurement using WLI. The temporal coherence of the fringes is then affected by the color filter at the pixels slightly decreasing vertical resolution of the measurement, and lateral sampling is half as fine as with monochromatic camera as only green pixels are used for the measurement, which are also the most sensitive pixels. Other interferometric systems with a color camera and white illumination are based on two- or three-wavelength (Red Green Blue) fringe analysis to calculate surface topography (Pförtner and Schwider 2003; Upputuri et al. 2014; Berwal et al. 2014; Ma et al. 2011; Guo et al. 2016) and even film thickness (Kitagawa 2014).

---

## Common Applications of WLI Systems

The development of WLI for surface topography measurement grew both as an extension of laser interferometry used for displacement measurement at a single point and as an extension of phase measurement with interference microscope objectives for topography measurement. Displacement measurement was based on a long scan at a single point, or maximally at a few points, while area interferometric topography measurement was based on processing only a few interferograms collected around the best focus using a phase-shifting interferometry method. Displacement methods started using white light interference, which allowed for distance measurement in, for example, the measurement of gauge blocks (Fluornoy et al. 1972). Single-point methods were often based on fiber optic sensors (Koch and Ulrich 1990; Rao Jackson 1996). The natural fringe contrast decrease in WLI due to the high numerical aperture of the objectives or the broad spectrum source inspired the extended range measurements of step-like surfaces also. Both these methods relied on two scans, one with laser source and the other with white light (Ikonen and Riski 1993; Ai and Caber 1994). At first, these methods were not thought capable of measuring rough surfaces. However, only a few years later and with the help of better computers, researchers realized that not only could the shape but also rough surfaces (where the signal does not look at all like fringes) could be measured with an automated and computerized full-field WLI scan.

## **WLI as Auxiliary Signal to Measurement with Monochromatic Illumination**

A laser interference signal was commonly used to measure gauge block heights up to 1000 mm or distance measurement, but these applications had limitations. In the early 1990s, one of the first applications of WLI was proposed (Ikonen and Riski 1993) that overcame fringe ambiguity in measurement. WLI was used as a reference signal and created in an auxiliary interferometric setup and then both beams, laser and white light, were introduced into the common measuring interferometer. Signals from laser and white light channels were simultaneously processed, and the exact length of the measured gauge block was obtained while the signals' maxima of laser and white light occurred coincidentally.

A very similar application of WLI was used recently in a more complicated holographic microscope system for high-precision topography measurements of micro-objects (Lizewski et al. 2014). Precise measurements would not have been possible without accurate in-focus plane determination provided by the additional white light interferometer and calibration sample. Both systems worked in sequence to find best position of the holographic objective with respect to the calibration sample, and then this information was used to establish the best positions for the measurement of the sample.

## **Roughness Measurement**

Roughness measurement has traditionally been performed on a stylus profiler, which measures the surface roughness along one profile. However, roughness along a single profile does not give a full image of surface roughness. For this reason, interferometric methods were developed for large area roughness measurement; however, because of the speckle effect created by light reflected off the rough surface, for a long time it was thought that accurate measurement was not possible. Indeed, the speckle effect creates a random effect in the phase of the fringes, but the coherence envelope is a good representation of a rough surface. For roughness measurement,  $20\times50\times$  objectives should be used as they collect most of the scattered light while still imaging a fairly large area.

With the development of areal topography measurements, surface engineering science is just now moving toward parameters for areal characterization of surfaces and tying these parameters to device functionality (Muralikrishnan and Raja 2009). These areal-surface parameters (*S* parameters for roughness and texture over the measured area) developed by Stout (2000) are standardized by the International Organization for Standardization (ISO) (ISO 2012) so as to enable a far greater understanding of surface finish characteristics. To characterize roughness, the underlying shape of the sample can be subtracted by using robust Gaussian fitting (Brinkmann et al. 2000; ISO 2010).

Surface roughness aerial *S* parameters can be divided into four groups: amplitude, spatial, functional, and hybrid parameters. Amplitude parameters are 3D extensions

of 2D  $R$  parameters and have limited use in surface characterization. Spatial parameters quantify the randomness, peaks, and directionality of the surface and determine the texture of different surface finishes. Functional parameters are useful in determining the bearing ratio calculations for the volume of the surface finish that would be removed during an ideal polishing process or fluid retention of the surface finish, among other. Hybrid parameters are related to slopes and curvatures of the roughness and used for evaluation of surface gloss or wear resistance.

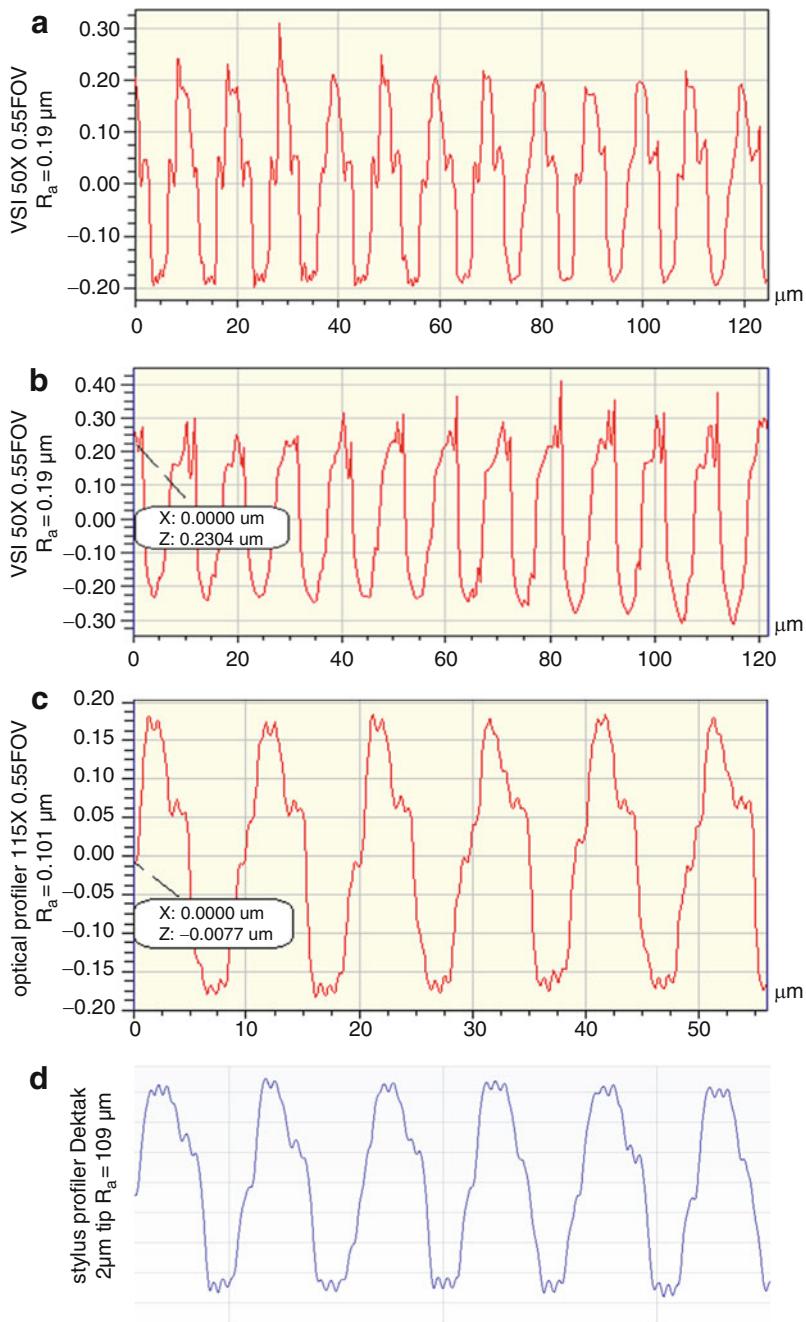
### Roughness Measurement Issues

However, while measuring roughness, caution needs to be taken as some microgeometry of the sample (described in section “[WLI Fringe Analysis to Measure Special Features of Objects](#)”), which include high slopes, V groove-type features, abrupt discontinuities, and randomness of surface, can affect the measurement. These microgeometries can cause the beam to undergo scattering, multiple reflections, diffraction, and interference, all of which affect the position and shape of the correlogram fringes.

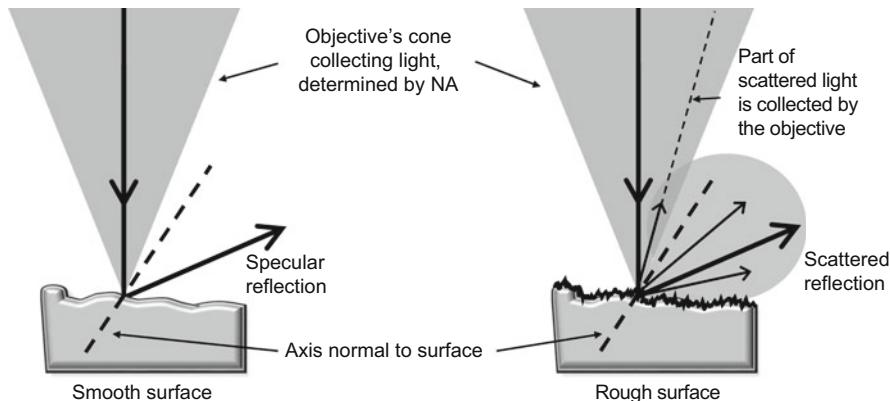
As pointed out in previous sections, surface roughness may consist of many different features in both character and size. Some surfaces may need to be measured with highest magnification (highest NA) for best collection of light scattered of the surface and finest sampling.

If a smaller magnification is chosen for the measurement, then a smaller portion of the light scattered from the sample is collected by the objective, causing loss of information and error in the measurement. Figure 19 shows the measurement of a near sinusoidal artifact structure with a 10- $\mu\text{m}$  period, which can be thought of as a representation of surface roughness. The two surface profiles were measured with the same  $50\times$  objective but used different fringe analysis algorithms; their results differ in shape and have clearly visible artificial spikes. In addition, the overall amplitude of the profile is higher than expected. The same sample was measured with higher numerical aperture and a  $115\times$  magnification objective, which collects more of the sample’s scattered light. All three results were compared with a measurement from a stylus profiler, which is well established in taking surface roughness measurements. This measurement confirms that the artifact is not perfectly sinusoidal and correlates well to the measurement with a  $115\times$  objective. Thus, the choice of the objective, algorithm, and system setup may have significant consequences on surface measurement. Some other efforts in identifying measurement error sources are reported in the literature ([Lehmann 2003](#); [Gao et al. 2008](#)), and some efforts to improve the roughness measurement, especially to minimize the speckle effect ([Wiesner et al. 2012](#)), were also suggested.

Despite some issues in roughness measurement, WLI is a very powerful tool for roughness evaluation. In addition, some roughness on the object actually helps when measuring objects that have higher slopes, as the rough surface scatters the light and some of the light is returned to the objective as shown in Fig. 20, which provides height information at many (but not all) points on the object’s slope. For objects with high slopes and smooth surfaces, that light is reflected outside of the objective and no height information can be determined.



**Fig. 19** Measurement results of near sinusoidal grating with (a) WLI with 115× objective (b) stylus profiler with 2- $\mu\text{m}$  tip (c) and (d) two different 50× objectives



**Fig. 20** Specular and diffusive reflection and light gathering cone of the objective

### Shape and Waviness Measurement

For shape and waviness measurement, a larger area of the object needs to be measured; thus, objectives of lower magnification,  $5\times$  and below, are used, but at the cost of loss of lateral resolution. If high resolution needs to be maintained, then slightly overlapping consecutive measurements can be taken and stitched together (Wyant and Schmit 1998).

### Examples of Applications

Many different areas of manufacturing rely on white light interferometry for shape and roughness measurement. Below we reference a few examples of WLI applications in different areas, for example, for surface characterization in material science using ion sputtering and laser ablation (Baryshev et al. 2012), in metal additive manufacturing (Gomez et al. 2017), in tribology for roughness measurement of ball bearings (Schmit et al. 2009), and for wear and corrosion measurements. WLI is also used to measure structured engineering surfaces like diffraction optics or patterned sapphire structures in LED manufacturing. In addition, softer materials, too delicate for using a stylus, are characterized using WLI. Paper in the printing industry, cloth and material manufacturing, and automotive paint finish are examples of this application; biology, for example for biofilm development (Brann et al. 2017) or contact lenses in ophthalmology (Giraldez et al. 2010) also employs WLI to measure delicate surfaces. Another unique application is using WLI to measure gunshot marks for gun identification (Rhee et al. 2006). These are only a few examples; new applications seem to appear every day.

Once the surface is measured, several analysis packages are available to measure different parameters like depth, width, diameter, spacings, angle, and statistical analysis.

## Other Systems Based on WLI

Many commercial systems for full-field noncontact surface topography measurement are based on scanning white light interferometry. However, in the labs, the stronger focus seems to be on the method that allows for the measurement of a single profile or film thickness from only a single frame of registered spectral fringes. This spectrally resolved white light interferometry method is briefly described in the next section.

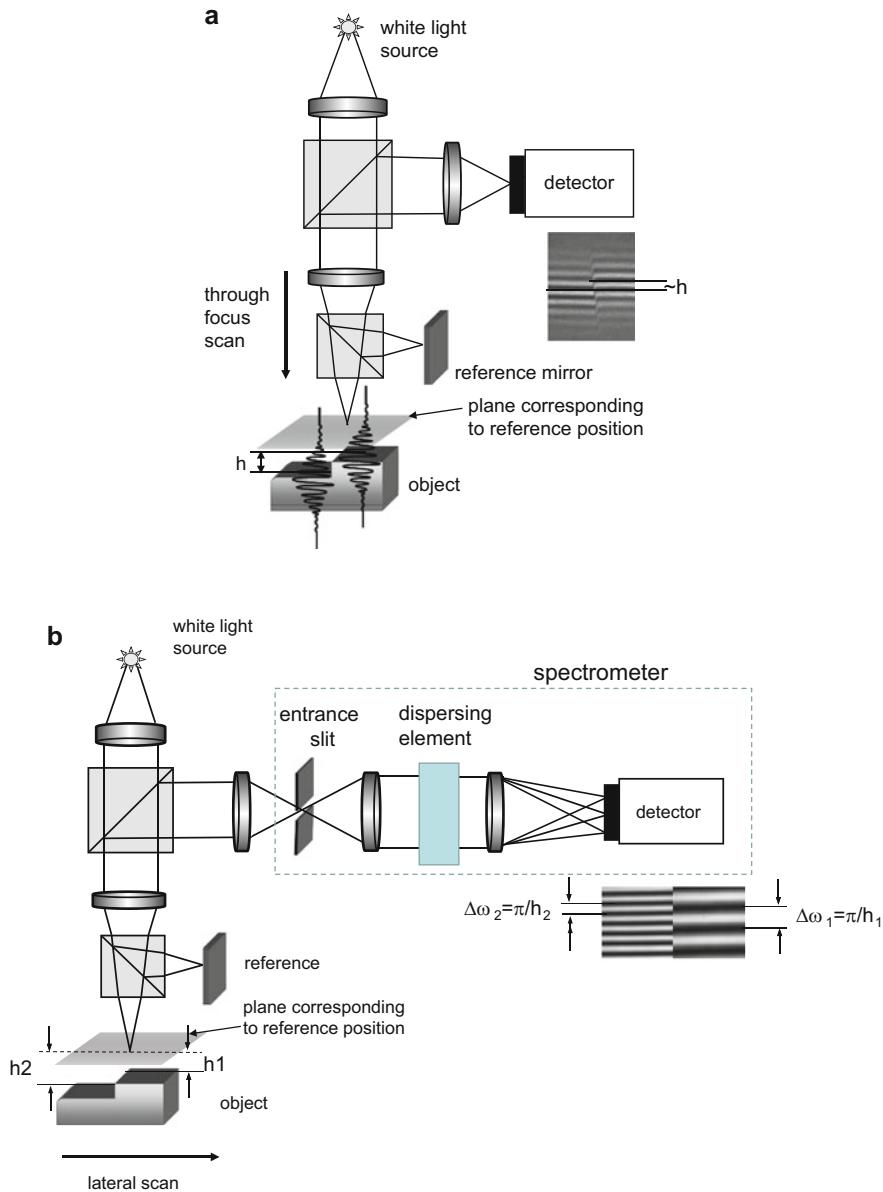
### Spectrally Resolved White Light Interferometry

Another system based on white light interference forms spectrally resolved fringes from a chosen single point or line of the interferogram. A similar technique is commonly used in optical coherence tomography (OCT) where the system scans an object point by point. In material sciences and engineering, measurement based on a line scan is preferred. In this method, the narrow slit of a spectrometer is placed at the image plane instead of the detector, as shown in Fig. 21. This method of white light interference observation is called spectrally resolved white light interferometry. Spectrally resolved white light interferometry is also called dispersive interferometry or multichannel chromatic interferometry. Although the first spectrally resolved white light interferometer for surface profilometry described by Schwider and Zhou (1994) was based on the Fizeau interferometer, these systems are typically based on the Michelson interferometers. The spectrometer spreads the single line of the object's white light fringes into spectral fringes along chromatic axis of solid state camera. Different types of spectrometers can be used, such as a prism or grating spectrometer.

The observed fringes are basically the “channeled spectrum” that scientists observed about 100 years ago. In this respect, spectrally resolved interferometry is also not a new technique; however, the development of computers and solid-state devices and spectrometers have allowed for the utilization of the channel spectrum for many purposes. Spectrally resolved white light fringes are used for many different applications, like measuring the differential index of refraction, the index of refraction distribution, wavelength multiplexing, transmission of images, distance and displacement measurements, and more recently profile measurement.

In the spectrally resolved white light interferometer, the chromatic axis is along one of the axes of camera, which is perpendicular to the slit of the spectrometer. The row (or column) of camera pixels registers a spectral interference signal from which the spectral phase or fringe frequency is calculated for a single object's point, delivering information about the object's distance from the corresponding reference point, and one camera frame delivers information about the profile of the object along one line. Thus, the object needs to be scanned laterally in order to obtain a 3D profile.

The spectral interference signal registered by each pixel delivers the information about the object's distance from the corresponding reference point, and collection of such information for all pixels creates the topographical map of the object.



**Fig. 21** Schematic of (a) white light and (b) spectrally resolved white light interferometer

The dynamic range for the measurement is limited by the spectral resolution of the spectrometer and the resolution is determined by the spectral bandwidth of the light source; thus, supercontinuum sources are preferred for systems requiring high resolution, but these systems become very expensive.

In order to determine the distance of the object with respect to the reference plane (equivalent to the best focus), a single spectral interferogram can be analyzed by measuring the frequency of the fringes or the spectral phase slope (Schwider and Zhou 1994). Both methods require calculating the carrier frequency of the fringes; the carrier frequency can be controlled by changing the distance of the objective to the object so that the object is just above or below focus. The slope of the spectral phase  $\Delta\varphi/2\Delta\omega$  can be measured using the Fourier transform (Takeda and Yamamoto 1994) or a spatial phase-shifting method (Sandoz et al. 1996). Because these spatial methods may be affected by intensity variations of the object across the field, using a temporal phase shifting technique was suggested (Debnath and Kothiyal 2005; Helen et al. 2001), where the PZT or achromatic phase shifter based on polarization elements would introduce the required phase shift so as to obtain a better phase and thus a better linear fit to determine the phase slope. By introducing a temporal phase shift, having a high carrier frequency of fringes is not required and the sample can be in or on either side of focus; this extends the range of the measurable sample height. However, because a few frames need to be collected using this method, measurement time slows considerably.

Spectrally resolved interferometry also can be used to measure films down to 2 μm by mapping the peaks of the Fourier transform; this method has been widely applied to biomedical sciences and engineering (Häusler and Lindner 1998; Zuluaga and Richards-Kortum 1999). However, film thickness down to 100 nm as well as the sample profile can be measured if the nonlinear portion of the spectral phase is analyzed (Debnath et al. 2006a, b) or reflectometry method alongside spectrally resolved interferometry is used (Debnath et al. 2009).

Instead of spectrally resolving white light fringes in space with a spectrometer, the source can be scanned through the range of wavelengths (wavelength scanning interferometry, WSI) and similar fringe analysis applied. WSI methods, due to the possibility of finer wavelength sampling, can be used for the measurement of larger thicknesses such as transparent optical plates.

It is worth pointing out again that there is a direct correspondence between interferograms collected by the types of interferometers described here, namely the scanning white light interferometer, the spectrally resolved white light interferometer, and the wavelength scanning interferometer. In each of these systems, the analysis can be based on the measurement of the wavelength-dependent OPD called also a spectral phase. For the spectrally resolved white light interferometer and wavelength scanning interferometer, the spectral decomposition of the interferogram is done by physical means, while in the white light interferometer the spectral decomposition is done mathematically. All these three methods have their equivalent methods in optical coherence tomography.

---

## WLI in Literature

White light interferometry is the most common name for interferometers based on white light used for surface measurement. The International Standard Organization chooses to call these methods low coherence interferometry (LCI) so as to include

different wavelength ranges and spatial effects that create a low coherence effect on fringe localization. However, in literature, LCI methods can be found under many different names; thus, it is important to know these other names when searching for information about WLI or LCI. Sometimes systems based on white light interface are called optical profilometers, which may suggest that a single profile is measured, which is the case in SRWLI but not with WLI systems that use an axial scan.

White light interferometry for the measurement of engineering surfaces has been variously labeled and can be found also under the following names:

White light interferometry (WLI)  
Vertical scanning interferometry (VSI)  
Low coherence interferometry (LCI)  
Coherence probe  
Optical coherence profilometry (OCP)  
Optical coherence microscopy  
Scanning white light interferometer (SWLI)  
White light scanning interferometry (WLSI)  
Coherence probe microscopy (CPM)  
Correlation microscopy  
Phase correlation microscopy  
Interference microscope  
Microscopic interferometry  
Wide band interferometry  
Full-field OCT  
Wide-field OCT  
Coherence radar  
Fringe peak scanning interferometry  
White Light Interference 3D microscope  
Rough surface tester

The WLI equivalent method for biological samples is most commonly called optical coherence tomography (OCT), but it also can be called time domain OCT (TD-OCT), coherence radar, or confocal interference microscopy.

The reader is encouraged to search for more information in the current and established literature (de Groot 2015; Schmit 2015) and articles, especially when looking for white light interferometry applications.

---

## References

- Abdulhalim I (2001) Theory for double beam interference microscopes with coherence effects and verification using the linnik microscope. *J Mod Opt* 48:279–302. <https://doi.org/10.1080/09500340108232458>
- Abdulhalim I (2006) Competence between spatial and temporal coherence in full field optical coherence tomography and interference microscopy. *J Opt A Pure Appl Opt* 8:952. <https://doi.org/10.1088/1464-4258/8/11/004>

- Abdulhalim I (2012) Spatial and temporal coherence effects in interference microscopy and full-field optical coherence tomography. *Ann Phys* 524:787–804. <https://doi.org/10.1080/09500340108232458>
- Ai C, Caber PJ (1994) Combination of white-light scanning and phase-shifting interferometry for surface profile measurements. US Patent US 5471303A. <https://patents.google.com/patent/US5471303A/en>. Accessed 20 Jun 2018
- Albertazzi AG Jr, Dal Pont A (2005) Fast coherence scanning interferometry for smooth, rough and spherical surfaces. *Fringe* 2005:605–612
- Albertazzi AG, Viotti MR, Miggiorin RM, Dal Pont A (2008) Applications of a white light interferometer for wear measurement of cylinders. *Proc SPIE* 7064:70640B. <https://doi.org/10.1117/12.796058>
- Balasubramanian N (1982) Optical system for surface topography measurement. US Patent 4,340,306
- Baryshev SV, Zinovev AV, Tripa CE, Erck RA, Veryovkin IV (2012) White light interferometry for quantitative surface characterization in ion sputtering experiments. *Appl Surf Sci* 258:6963–6968. <https://doi.org/10.1016/j.apsusc.2012.03.144>
- Bauer W (2003) Special properties of coherence scanning interferometers for large measurement volumes. *J Phys Conf Ser* 311:012030. <https://doi.org/10.1088/1742-6596/311/1/012030>
- Beverage JL, de Lega XC, Fay MF (2014) Interferometric microscope with true color imaging. *SPIE Proc* 9203:92030S. <https://doi.org/10.1117/12.2063264>
- Berwal S, Srivastava V, Dubey VK, Ahmad A, Mehta DS, md. Inam (2014) High-resolution white light interferometry for quantitative phase imaging of human red blood cells using three-chip colour camera. International conference on fibre optics and photonics:M4A.6. <https://doi.org/10.1364/PHOTONICS.2014.M4A.6>
- Bosseboeuf A, Petitgrand S (2003) Characterization of the static and dynamic behaviour of M(O) EMS by optical techniques: status and trends. *J Micromech Microeng* 13:S23–S33. <https://doi.org/10.1088/0960-1317/13/4/304>
- Bouchal P, Čelechovský R, Bouchal Z (2015) Polarization sensitive phase-shifting Mirau interferometry using a liquid crystal variable retarder. *Opt Lett* 40:4567. <https://doi.org/10.1364/OL.40.004567>
- Brann M, Suter JD, Addleman RS, Larimer C (2017) Monitoring bacterial biofilms with a microfluidic flow chip designed for imaging with white-light interferometry. *Biomicrofluidics* 11:044113. <https://doi.org/10.1063/1.4985773>
- Brinkmann S, Bodschwinna H, Lemke H-W (2000) Development of robust Gaussian regression filter for three dimensional surface analysis. Proceedings of the 10th international colloquium on surfaces, Chemnitz, Germany
- Cai Y, Dong Y, Hoenders BJ (2012) Interdependence between the temporal and spatial longitudinal and transverse degrees of partial coherence and a generalization of the van Cittert–Zernike theorem. *J Opt Soc Am A* 29:2542–2551. <https://doi.org/10.1364/JOSAA.29.002542>
- Caber PJ (1993) Interferometric profiler for rough surfaces. *Appl Opt* 32:3438–3441. <https://doi.org/10.1364/AO.32.003438>
- Chen D (2006) High-definition vertical-scan interferometry. US Patent US7605925B1
- Chen L-C, Huang Y-T, Nguyen X-L, Chen J-L, Chang C-C (2009) Dynamic out-of-plane profilometry for nano-scale full-field characterization of MEMS using stroboscopic interferometry with novel signal deconvolution algorithm. *Opt Laser Eng* 47:237–251. <https://doi.org/10.1016/j.optlaseng.2008.05.016>
- Creath K (1997) Sampling requirements for white light interferometry. *Fringe'97* proceedings of the 3rd international workshop on automatic processing of fringe patterns, pp 52–59
- Davidson M, Kaufman K, Mazor I, Cohen F (1987) An application of interference microscopy to integrated circuit inspection and metrology. *Proc SPIE* 0775. <https://doi.org/10.1117/12.940433>
- Debnath SK, Kothiyal MP (2005) Optical profiler based on spectrally resolved white light interferometry. *Opt Eng* 44:013606. <https://doi.org/10.1117/1.1828468>

- Debnath SK, Kothiyal MP, Schmit J, Hariharan P (2006a) Spectrally resolved white-light phase-shifting interference microscopy for thickness-profile measurements of transparent thin film layers on patterned substrates. *Opt Express* 14:4662–4667. <https://doi.org/10.1364/OE.14.004662>
- Debnath SK, Kothiyal MP, Schmit J, Hariharan P (2006b) Spectrally resolved phase-shifting interferometry of transparent thin films: sensitivity of thickness measurements. *Appl Opt* 45:8636–8640. <https://doi.org/10.1364/AO.45.008636>
- Debnath SK, You J, Kim S-W (2009) Determination of film thickness and surface profile using reflectometry and spectrally resolved phase shifting interferometry. *Int J Precis Eng Manuf* 10 (5–10). <https://doi.org/10.1007/s12541-009-0086-0>
- Deck L, de Groot P (1994) High-speed noncontact profiler based on scanning white-light interferometry. *Appl Opt* 33:7334–7338. <https://doi.org/10.1364/AO.33.007334>
- Dong JLR (2012) Sensitivity analysis of thin-film thickness measurement by vertical scanning white-light interferometry. *Appl Opt* 51:5668–5675. <https://doi.org/10.1364/AO.51.005668>
- Farr K, George N (1992) Beamsplitter cube for white light interferometry. *Opt Eng* 31:2191–2196. <https://doi.org/10.1117/12.59886>
- Fay MF, Dresel T (2017) Applications of model-based transparent surface films analysis using coherence-scanning interferometry. *Opt Eng* 56:111709. <https://doi.org/10.1117/12.2238997>
- Flournoy PA, McClure RW, Wyntjes G (1972) White-light interferometric thickness gauge. *Appl Opt* 11:1907–1915. <https://doi.org/10.1364/AO.11.001907>
- Gao F, Coupland JM, Petzing J (2006) V-groove measurements using white light interferometry. *Photon06*, Manchester
- Gao F, Leach RK, Petzing J, Coupland J (2008) Surface measurement errors using commercial scanning white light interferometers. *Meas Sci Technol* 19:015303. <https://doi.org/10.1088/0957-0233/19/1/015303>
- Giraldez MJ, García-Resúa C, Lira M, Real Oliveira ME, Yebra-Pimentel E (2010) White light interferometry to characterize the hydrogel contact lens surface. *Ophthalmic Physiol Opt* 30:289–297. <https://doi.org/10.1111/j.1475-1313.201000712x>
- Gokhler M, Rosen J (2005a) General configuration for using the longitudinal spatial coherence effect. *Opt Commun* 252:22–28. <https://doi.org/10.1016/j.optcom.2005.04.007>
- Gokhler M, Rosen J (2005b) Synthesis of a multiple-peak spatial degree of coherence for imaging through absorbing media. *Appl Opt* 44:2921–2927. <https://doi.org/10.1364/AO.44.002921>
- Gomez C, Su R, Thompson A, DiSciacca J, Lawes S, Leach RK (2017) Optimization of surface measurement for metal additive manufacturing using coherence scanning interferometry. *Opt Eng* 56:111714. <https://doi.org/10.1117/1.OE.56.11.111714>
- de Groot P, Deck L (1993) Three-dimensional imaging by sub-Nyquist sampling of white-light interferograms. *Opt Lett* 18:1462–1464. <https://doi.org/10.1364/OL.18.001462>
- de Groot P, Deck L (1995) Surface profiling by analysis of white-light Interferograms in the spatial frequency domain. *J Mod Opt* 42:389–401. <https://doi.org/10.1080/09500349514550341>
- de Groot P, de Lega XC, Kramer J, Turzhitsky M (2002) Determination of fringe order in white-light interference microscopy. *Appl Opt* 41:4571–4578. <https://doi.org/10.1364/AO.41.004571>
- de Groot P, de Lega XC (2003) Valve cone measurement using white light interference microscopy in a spherical measurement geometry. *Opt Eng* 42:1232–1237. <https://doi.org/10.1117/1.1565350>
- de Groot P, de Lega XC (2004) Signal modeling for low-coherence height-scanning interference microscopy. *Appl Opt* 43:4821–4830. <https://doi.org/10.1364/AO.43.004821>
- de Groot P (2005) Generating fringe-free images from phase-shifted interferometry data. *Appl Opt* 44:7062–7069. <https://doi.org/10.1364/AO.44.007062>
- de Groot P (2006) Stroboscopic white-light interference microscopy. *Appl Opt* 45:5840–5844. <https://doi.org/10.1364/AO.45.005840>
- de Groot P, de Lega XC (2007a) Angle-resolved three-dimensional analysis of surface films by coherence scanning interferometry. *Opt Lett* 32:1638–1640. <https://doi.org/10.1364/OL.32.001638>

- de Groot P, de Lega X (2007b) Triangulation methods and systems for profiling surfaces through a thin film coating. US Patent 7,292,346, 2007. <https://patents.google.com/patent/US7292346B2/en>. Accessed 19 Jun 2018
- de Groot P (2008) Method and system for analyzing low-coherence interferometry signals for information about thin film structures. US Patent 7,321,431. <https://patents.google.com/patent/US20060262321>. Accessed 12 Jun 2018
- de Groot P, de Lega X, Liesener J, Darwin M (2008) Metrology of optically-unresolved features using interferometric surface profiling and RCWA modeling. Opt Express 16:3970–3975. <https://doi.org/10.1364/OE.16.003970>
- de Groot P, Biegen J (2015) A new class of wide-field objectives for 3D interference microscopy. Proc SPIE 9525:95250N. <https://doi.org/10.1117/12.2183628>
- de Groot P (2015) Interference microscopy for surface structure analysis. In: Yoshizawa T (ed) Handbook of optical metrology principles and applications, second edition. CRC Press, Boca Raton, pp 791–828
- de Groot P (2016) The meaning and measure of vertical resolution in surface metrology. 5th international conference on surface metrology, Poznań, Poland, pp 1–4
- Guo T, Li F, Chen J, Fu X, Hu X (2016) Multi-wavelength phase-shifting interferometry for microstructures measurement based on color image processing in white light interference. Opt Lasers Eng 82:41–47. <https://doi.org/10.1016/j.optlaseng.2016.02.003>
- Han S, Novak E, Wissinger J, Guenther BW, Browne T, Yanine E, Schurig M, Herron JD, McCloy C, Li X, Krell MB, Harris J (2005) Surface profiler for fixed through-glass measurement. Proc SPIE 5716:189–197. <https://doi.org/10.1117/12.597080>
- Hanhijärvi K, Kassamakov I, Heikkilä V, Aaltonen J, Sainiemi L, Grigoras K, Franssila S, Hæggström E (2012) Stroboscopic supercontinuum white-light interferometer for MEMS characterization. Opt Lett 37:1703–1705. <https://doi.org/10.1364/OL.37.001703>
- Hanhijärvi K, Kassamakov I, Aaltonen J, Heikkilä V, Sainiemi L, Franssila S, Hæggström E (2013) Through-silicon stroboscopic characterization of an oscillating MEMS thermal actuator using supercontinuum interferometry. IEEE/ASME Trans Mechatron 18:1418–1420. <https://doi.org/10.1109/TMECH.2012.2235080>
- Harasaki A, Schmit J, Wyant JC (2000) Improved vertical-scanning interferometry. Appl Opt 39:2107–2115. <https://doi.org/10.1364/AO.39.002107>
- Harasaki A, Wyant JC (2000) Fringe modulation skewing effect in white-light vertical scanning interferometry. Appl Opt 39:2101–2106. <https://doi.org/10.1364/AO.39.002101>
- Harasaki A, Schmit J, Wyant JC (2001) Offset of coherent envelope position due to phase change on reflection. Appl Opt 40:2102–2106. <https://doi.org/10.1364/AO.40.002102>
- Hariharan P, Oreb BF, Eiju T (1987) Digital phase-shifting interferometry: a simple error-compensating phase calculation algorithm. Appl Opt 26:2504–2506. <https://doi.org/10.1364/AO.26.002504>
- Hariharan P (1995) Achromatic retarders using quartz and mica. Meas Sci Technol 6:1078–1079. <https://doi.org/10.1088/0957-0233/6/7/036>
- Hariharan P (1996) Achromatic phase-shifting for white-light interferometry. Appl Opt 35:6823–6824. <https://doi.org/10.1364/AO.35.006823>
- Hariharan P, Roy M (1996) Interferometric surface profiling with white light: effects of surface films. J Mod Opt 43:1797–1800. <https://doi.org/10.1080/09500349608232850>
- Haruna M, Ohmi M, Mitsuyama T, Tajiri H, Maruyama H, Hashimoto M (1998) Simultaneous measurement of the phase and group indices and the thickness of transparent plates by low-coherence interferometry. Opt Lett 23:966–968. <https://doi.org/10.1364/OL.23.000966>
- Häusler G, Herrmann JM (1993) Physical limits of 3D sensing. Proc SPIE 1822. <https://doi.org/10.1117/12.141937>
- Häusler G, Neumann J (1993) Coherence radar: an accurate 3D sensor for rough surfaces. Proc SPIE 1822. <https://doi.org/10.1117/12.141943>
- Häusler G, Lindner MW (1998) “Coherence radar” and “spectral radar”—new tools for dermatological diagnosis. J Biomed Opt 3:21–31. <https://doi.org/10.1117/1.429899>

- Heikkinen V, Kassamakov I, Paulin T, Nolvi A, Hæggström E (2013) Stroboscopic scanning white light interferometry at 2.7 MHz with 1.6  $\mu\text{m}$  coherence length using a non-phosphor LED source. *Opt Express* 21:5247–5254. <https://doi.org/10.1364/OE.21.005247>
- Heikkinen V, Kassamakov I, Seppä J, Paulin T, Nolvi A, Lassila A, Hæggström E (2015) Traceable methods for vertical scale characterization of dynamic stroboscopic scanning white-light interferometer measurements. *Opt Express* 54:10397–10403. <https://doi.org/10.1364/OE.54.010397>
- Helen SS, Kothiyal MP, Sirohi RS (2001) Analysis of spectrally resolved white light Interferograms: use of a phase shifting technique. *Opt Eng* 40:1329–1336. <https://doi.org/10.1117/1.1383997>
- Hirabayashi A, Ogawa H, Kitagawa K (2002) Fast surface profiler by white-light interferometry by use of a new algorithm based on sampling theory. *Appl Opt* 41:4876–4883. <https://doi.org/10.1364/AO.41.004876>
- Hlubina P, Luňáček J, Ciprian D, Chlebus R (2008) Dispersion error of a beam splitter cube in white-light spectral interferometry. *Opto-Electron Rev* 16:439–443. <https://doi.org/10.2478/s11772-008-0040-1>
- Hu X, Deng S, Wang W, Hu X (2009) Long-working-distance Linnik microscopic interferometry for measuring the surface profile of microstructures. *J Vac Sci Technol B* 27:1403. <https://doi.org/10.1116/1.3046154>
- Ikonen E, Riski K (1993) Gauge block interferometer based on one stabilised laser and a whitelight source. *Metrologia* 30:95–104. <https://doi.org/10.1088/0026-1394/30/2/006>
- ISO – International Organization for Standardization (2010) ISO/CD TS 16610-31:2010, geometrical product specifications (GPS) – filtration – part 31: robust profile filters: Gaussian regression filters. <https://www.iso.org/standard/42972.html>. Accessed 18 Jun 2018
- ISO – International Organization for Standardization (2012) ISO/DIS 25178-2:2012 geometrical product specification (GPS) – surface texture: areal – part 2 terms. Definitions and surface texture parameters. <https://www.iso.org/standard/42785.html>. Accessed 18 Jun 2018
- Kang M-G, Lee S-Y, Kim S-W (1999) Self-compensation of PZT errors in white light scanning interferometry. *J Opt Soc Korea* 3:35–40
- Kassamakov I (2008) Stroboscopic white light interferometry for dynamic characterization of capacitive pressure sensors. *J Acoust Soc Am* 123:3286. <https://doi.org/10.1121/1.2933663>
- Kim S, Kim G (1999) Thickness-profile measurement of transparent thin-film layers by white-light scanning interferometry. *Appl Opt* 38:5968–5973. <https://doi.org/10.1364/AO.38.005968>
- Kim J, Kim K, Pahk HJ (2017) Thickness measurement of a transparent thin film using phase change in white-light phase-shift interferometry. *Curr Opt Photon* 1:505–513
- Kimborough B, Millerd J, Wyant J, Hayes J (2006) Low coherence vibration insensitive Fizeau interferometer. *Proc SPIE* 6292:62920F. <https://doi.org/10.1117/12.682956>
- Kino GS, Chim SSC (1990) Mirau correlation microscope. *Appl Opt* 29:3775–3783. <https://doi.org/10.1364/AO.29.003775>
- Kitagawa K (2014) Surface and thickness profile measurement of a transparent film by three-wavelength vertical scanning interferometry. *Opt Lett* 39:4172–4175. <https://doi.org/10.1364/OL.39.004172>
- Koch A, Ulrich R (1990) Fiber-optic displacement sensor with 0.02  $\mu\text{m}$  resolution by white-light interferometry. *Sens. Actuators A Phys* 201–207(1990):25. [https://doi.org/10.1016/0924-4247\(90\)87032-E](https://doi.org/10.1016/0924-4247(90)87032-E)
- Koudelka PD, Burge JH (2004) Fabrication of cube beamsplitter for white light interferometry. *Proc SPIE* 5252:17–25. <https://doi.org/10.1117/12.513440>
- Kujawińska M, Józwik M, Styk A, Zeitner U, Roysten A, Beer S, Moosburger R, Gorecki C, Gastering K (2010) Design, technology and signal processing for DOE-based micro-interferometer array applied in new generation M(O)EMS test equipment. *Proc SPIE* 7791:779103–779101. <https://doi.org/10.1117/12.862170>
- Küchel MF (2004) Zygo's new method to reduce intrinsic noise in interferometers. <https://www.zygo.com/library/papers/optatec2004.pdf>. Accessed 13 Jun 2018
- Larkin KG (1996a) Efficient nonlinear algorithm for envelope detection in white light interferometry. *JOSA* 13:832–843. <https://doi.org/10.1364/JOSAA.13.000832>

- Larkin KG (1996b) Efficient demodulator for bandpass sampled AM signals. *Electron Lett* 32:101–102. <https://doi.org/10.1049/el:19960081>
- Lee BS, Strand TC (1990) Profilometry with a coherence scanning microscope. *Appl Opt* 29:3784–3788. <https://doi.org/10.1364/AO.29.003784>
- de Lega XC, de Groot P (2008a) Characterization of materials and film stacks for accurate surface topography measurement using a white-light optical profiler. *Proc SPIE* 6995:69950P. <https://doi.org/10.1117/12.782836>
- de Lega XC, de Groot P (2008b) Transparent film profiling and analysis by interference microscopy. *Proc SPIE* 7064:70640I. <https://doi.org/10.1117/12.794936>
- Lehmann P (2003) Optical versus tactile geometry measurement – alternatives or counterparts. *Proc SPIE* 5144:183–196. <https://doi.org/10.1117/12.508360>
- Li Y, Kästner M, Reithmeier E (2017) Vibration-insensitive low coherence interferometer (LCI) for the measurement of technical surfaces. *Measurement* 104:36–42. <https://doi.org/10.1016/j.optlaseng.2016.02.003>
- Lindner MW (2002) White-light interferometry via an endoscope. *Proc SPIE* 4777. <https://doi.org/10.1117/12.472208>
- Lindner M, Thominet V (2005) Optical measuring device. US Patent US 6970253 2005. <https://patents.google.com/patent/US6970253>. Accessed 20 Jun 2018
- Linnik VP (1933) Ein apparat fur mikroskopisch-interferometrische untersuchung reflektierender objekte (mikrointerferometer). *Akademija Nauk SSSR Doklady* 21:18–23
- Liżewski K, Tomczewski S, Kozacki T, Kostencka J (2014) High-precision topography measurement through accurate in-focus plane detection with hybrid digital holographic microscope and white light interferometer module. *Appl Opt* 53:2446–2454. <https://doi.org/10.1364/AO.53.002446>
- Lyakin DV, Ryabukho VP (2013) Longitudinal correlation properties of an optical field with broad angular and frequency spectra and their manifestation in interference microscopy. *Quantum Electron* 43:949. <https://doi.org/10.1070/QE2013v04n10ABEH015187>
- Lyulko OV, Randers-Pehrson G, Brenner DJ (2013) Simultaneous immersion Mirau interferometry. *Rev Sci Instrum* 84:053701. <https://doi.org/10.1063/1.4803181>
- Ma S, Quan C, Zhu R, Tay CJ, Chen L (2011) Surface profile measurement in white-light scanning interferometry using a three-chip color CCD. *Appl Opt* 50:2246–2254. <https://doi.org/10.1364/AO.50.002246>
- Machleidt T, Kollhoff D, Dathe O, Kapusi O, Nestler R, Franke K-H (2012) Application of color cameras for 3D surface measurements with white light interferometers. 14th SpectroNet collaboration forum AM in Jena
- Malacara D (2007) Optical shop testing, 3rd edn. Wiley, Hoboken
- Mansfield D (2006) The distorted helix: thin film extraction from scanning white light interferometry. *Proc SPIE* 6186:61860O. <https://doi.org/10.1117/12.664036>
- Mandel L, Wolf E (1995) Optical coherence and quantum optics, 1st edn. Cambridge University Press, Cambridge
- Mirau AH (1949) Interferometer. US Patent 2,612,074. <https://patents.google.com/patent/US2612074>. Accessed 12 Jun 2018
- Mun JI, Jo T, Kim T, Pahk HJ (2015) Residual vibration reduction of white-light scanning interferometry by input shaping. *Opt Express* 23:464–470. <https://doi.org/10.1364/OE.23.000464>
- Muralikrishnan B, Raja J (2009) Computational surface and roundness metrology. Springer-Verlag, London
- Niehues J, Lehmann P, Xie W (2012) Low coherent Linnik interferometer optimized for use in nano-measuring machines. *Meas Sci Technol* 23:125002. <https://doi.org/10.1088/0957-0233/23/12/125002>
- Novotny S, Durairaj V, Shavrin I, Lipiäinen L, Kokkonen K, Kaivola M, Ludvigsen H (2014) Picosecond supercontinuum light source for stroboscopic white-light interferometry with freely adjustable pulse repetition rate. *Opt Express* 22:13625–13633. <https://doi.org/10.1364/OE.22.013625>
- Olszak A, Schmit J (2003) High-stability white-light interferometry with reference signal for real-time correction of scanning errors. *Opt Eng* 42:54–59. <https://doi.org/10.1117/1.1523942>

- Palmer JM (1993) Getting intense about intensity. *Metrologia* 30:371–372. <https://doi.org/10.1088/0026-1394/30/4/030>
- Park M-C, Kim S-W (2001) Compensation of phase change on reflection in white-light interferometry for step height measurement. *Opt Lett* 26:420–422. <https://doi.org/10.1364/OL.26.000420>
- Pavliček P, Soubusta J (2003) Theoretical measurement uncertainty of white-light interferometry on rough surfaces. *Appl Opt* 42:1809–1813. <https://doi.org/10.1364/AO.42.001809>
- Pavliček P, Soubusta J (2004) Measurement of the influence of dispersion on white-light interferometry. *Appl Opt* 43:766–770. <https://doi.org/10.1364/AO.43.000766>
- Pavliček P, Halouzka M, Duan Z, Takeda M (2009) Spatial coherence profilometry on tilted surfaces. *Appl Opt* 48:H40–H47. <https://doi.org/10.1364/AO.48.000H40>
- Petitgrand S, Yahiaoui R, Danaie K, Bosseboeuf A, Gilles J (2001) 3D measurement of micro-mechanical devices vibration mode shapes with a stroboscopic interferometric microscope. *Opt Laser Eng* 36:77–101. [https://doi.org/10.1016/S0143-8166\(01\)00040-9](https://doi.org/10.1016/S0143-8166(01)00040-9)
- Pfärtner A, Schwider J (2001) Dispersion error in white-light Linnik interferometers and its applications for evaluation procedures. *Appl Opt* 40(34):6223–6228. <https://doi.org/10.1364/AO.40.006223>
- Pförtner A, Schwider J (2003) Red-green-blue interferometer for the metrology of discontinuous structures. *Appl Opt* 42:667–673. <https://doi.org/10.1364/AO.42.000667>
- Rao Y-J, Jackson DA (1996) Recent progress in fibre optic low-coherence interferometry. *Meas Sci Technol* 7:981–999. <https://doi.org/10.1088/0957-0233/7/7/001>
- Recknagel R, Notni G (1998) Analysis of white light interferograms using wavelet methods. *Opt Commun* 148:122–128. [https://doi.org/10.1016/S0030-4018\(97\)00644-5](https://doi.org/10.1016/S0030-4018(97)00644-5)
- Reolon D, Jacquot M, Verrier I, Brun G, Veillas C (2006) Broadband supercontinuum interferometer for high-resolution profilometry. *Opt Express* 14:128–137. <https://doi.org/10.1364/OPEX.14.000128>
- Rhee H-G, Lee Y-W, Vorburger TV, Reneger TB (2006) Surface topography measurement and analysis for bullet and casing signature identification. *Korean J Opt Photon* 17:47–53. <https://doi.org/10.3807/KJOP.2006.17.1.047>
- Rosen J, Takeda M (2000) Longitudinal spatial coherence applied for surface profilometry. *Appl Opt* 39:4107–4111. <https://doi.org/10.1364/AO.39.004107>
- Roy M, Cooper I, Moore P, Sheppard CJR, Hariharan P (2005) White-light interference microscopy: effects of multiple reflections within a surface film. *Opt Express* 13:164–170. <https://doi.org/10.1364/OPEX.13.000164>
- Ryabukho V, Lyakin D, Lobachev M (2004) Influence of longitudinal spatial coherence on the signal of a scanning interferometer. *Opt Lett* 29:667–669. <https://doi.org/10.1364/OL.29.000667>
- Safrani A, Abdulhalim I (2011) Spatial coherence effect on layer thickness determination in narrowband full-field optical coherence tomography. *Appl Opt* 50:3021–3027. <https://doi.org/10.1364/AO.50.003021>
- Safrani A, Abdulhalim I (2012) Ultrahigh-resolution full-field optical coherence tomography using spatial coherence gating and quasi-monochromatic illumination. *Opt Lett* 37:458–460. <https://doi.org/10.1364/OL.37.000458>
- Saleh BEA, Teich MC (2007) Fundamentals of photonics, 2nd edn. Wiley, Hoboken
- Salsbury C, Olszak A (2017) Spectrally controlled interferometry. *Appl Opt* 56:7781–7788. <https://doi.org/10.1364/AO.56.007781>
- Sandoz P, Tribillon G, Perrin H (1996) High resolution profilometry by using phase calculation algorithms for spectroscopic analysis of white light Interferograms. *J Mod Opt* 43:701–708. <https://doi.org/10.1080/09500349608232777>
- Sandoz P (1997) Wavelet transform as a processing tool in white-light interferometry. *Opt Lett* 22:1065–1067. <https://doi.org/10.1364/OL.22.001065>
- Schmit J, Olszak A (2002) High-precision shape measurement by white-light interferometry with real-time scanner error correction. *Appl Opt* 41:5943–5950. <https://doi.org/10.1364/AO.41.005943>

- Schmit J (2003) High speed measurements using optical profiler. Proc SPIE 5144:46–56. <https://doi.org/10.1117/12.501035>
- Schmit J, Unruh P, Tavrov AV, Kerwien N, Osten W, Tiziani H (2003) Coherence effects in narrow trench measurements using white light interferometry. Frontiers in Optics, OSA Technical Digest, MT41. <https://doi.org/10.1364/FIO.2003.MT41>
- Schmit J, Hariharan P (2007) Improved polarization Mirau interference microscope. Opt Eng 46:077007. <https://doi.org/10.1117/1.2751166>
- Schmit J, Han S, Novak E (2009) Ball bearing measurement with white light interferometry. Proc SPIE 7389:73890P. <https://doi.org/10.1117/12.828113>
- Schmit J (2013) White-light interference 3D microscopes. In: Harding K (ed) Handbook of optical dimensional metrology. Taylor & Francis, Boca Raton, pp 395–418
- Schmit J, Novak M, Bui S (2015) 3D interferometric microscope: color visualization of engineered surfaces for industrial applications. Proc SPIE 9576:957607. <https://doi.org/10.1117/12.2190352>
- Schmit J (2015) 3D white light interference microscope with specialized illumination for better sample imaging and observation. In: Jabłoński R, Brezina T (eds) Advanced mechatronics solutions. Springer, Cham, pp 553–558
- Schreiber H, Bruning JH (2007) Phase shifting interferometry. Ch. In: Malacara D (ed) Optical shop testing., 3rd edn, vol 14. Wiley, Hoboken, pp 547–666
- Schulz LG (1951) The effect of phase changes in white light interferometry. J Opt Soc Am 41:261–264. <https://doi.org/10.1364/JOSA.41.000261>
- Schwider J, Burow R, Elssner K-E, Grzanna J, Spolarczyk R, Merkel K (1983) Digital wave-front measuring interferometry: some systematic error sources. Appl Opt 22:3421–3432. <https://doi.org/10.1364/AO.22.003421>
- Schwider J, Zhou L (1994) Dispersive interferometric profilometer. Opt Lett 19:995–997. <https://doi.org/10.1364/OL.19.000995>
- Shavrin I, Lipiäinen L, Kokkonen K, Novotny S, Kaivola M, Ludvigsen H (2013) Stroboscopic white-light interferometry of vibrating microstructures. Opt Express 21:16901–16907. <https://doi.org/10.1364/OE.21.016901>
- Stout KJ (ed) (2000) Development of methods for the characterisation of roughness in three dimension. Penton Press, London
- Styk A, Lambelet P, Royset A, Kujawińska M, Gastinger K (2010) Smart pixel camera based singal processing in an interferometric test stadion for massive paralel inspection of MEMS and MOEMS. Proc SPIE 7387:73870M-1/10. <https://doi.org/10.1117/12.871522>
- Sun C, Yu L, Sun Y, Yu Q (2005) Scanning white-light interferometer for measurement of the thickness of a transparent oil film on water. Appl Opt 44:5202–5205. <https://doi.org/10.1364/AO.44.005202>
- Takeda M, Yamamoto H (1994) Fourier-transform speckle Profilometry: three-dimensional shape measurements of diffuse objects with large height steps and/or spatially isolated surfaces. Appl Opt 33:7829–7837. <https://doi.org/10.1364/AO.33.007829>
- Tapilouw AM, Chen LC, Jen YJ, Lin ST, Yeh SL (2013) Orthogonal polarization Mirau interferometer using reflective-type waveplate. Opt Lett 38:2502–2504. <https://doi.org/10.1364/OL.38.002502>
- Tavrov A, Schmit J, Kerwien N, Osten W, Tiziani H (2005) Diffraction-induced coherence levels. Appl Opt 44:2202–2212. <https://doi.org/10.1364/AO.44.002202>
- Tereschenko S, Lehmann P, Zellmer L, Brueckner-Foit A (2016) Passive vibration compensation in scanning white-light interferometry. Appl Opt 55:6172–6182. <https://doi.org/10.1364/AO.55.006172>
- Tomczewski S, Pakuła A, Van Erps J, Thienpont H, Salbut L (2013) Low-coherence interferometry with polynomial interpolation on compute unified device architecture-enabled graphics processing units. Opt Eng 52:094105. <https://doi.org/10.1117/1.OE.52.9.094105>
- Upputuri PK, Gong L, Wang H, Pramanik M, Mohan Nandigana K, Kothiyal MP (2014) Measurement of large discontinuities using single white light interferogram. Opt Express 22:27373–27380. <https://doi.org/10.1364/OE.22.027373>

- Wan DS (2009) Measurement of thin films using Fourier amplitude. US Patent 7,612,891. <https://patents.google.com/patent/US7612891B2/en>. Accessed 12 Jun 2018
- Wang W, Kozaki H, Rosen J, Takeda M (2002) Synthesis of longitudinal coherence functions by spatial modulation of an extended light source: a new interpretation and experimental verifications. *Appl Opt* 41:1962–1971. <https://doi.org/10.1364/AO.41.001962>
- Wang D, Liang R (2016) Simultaneous polarization Mirau interferometer based on pixelated polarization camera. *Opt Lett* 41:41–44. <https://doi.org/10.1364/OL.41.000041>
- Wiersma JT, Wyant JC (2013) Vibration insensitive extended range interference microscopy. *Appl Opt* 52:5957–5961. <https://doi.org/10.1364/AO.52.005957>
- Wiesner B, Hybl O, Häusler G (2012) Improved white-light interferometry on rough surfaces by statistically independent speckle patterns. *Appl Opt* 51:751–757. <https://doi.org/10.1364/AO.51.000751>
- Wyant JC, Schmit J (1998) Large field of view, high spatial resolution, surface measurements. *Int J Mach Tools Manufact* 38:691–698
- Xu Z, Shilpiekandula V, Youcefoumi K, Yoon SF (2009) White-light scanning interferometer for absolute nano-scale gap thickness measurement. *Opt Express* 17:15104–15117. <https://doi.org/10.1364/OE.17.015104>
- Yamada R., Itoh M, Tian RL, Tsai M, Yatagai T (1994) Broad-band light-wave correlation topography using wavelet transform. Technical abstract, ICO topical meeting, Kyoto
- Yatagai T (1994) Recent progress in white light interferometry. *Proc SPIE* 2340:338–345. <https://doi.org/10.1117/12.195930>
- Yoshino H, Abbas A, Kaminski PM, Smith R, Walls JM, Mansfield D (2017) Measurement of thin film interfacial surface roughness by coherence scanning interferometry. *J Appl Phys* 121:105303. <https://doi.org/10.1063/1.4978066>
- Zuluaga AF, Richards-Kortum R (1999) Spatially resolved spectral interferometry for determination of subsurface structure. *Opt Lett* 24:519–521. <https://doi.org/10.1364/OL.24.000519>



# Optical Coherence Tomography for NDE

15

Jonas Golde, Lars Kirsten, Christian Schnabel, Julia Walther, and Edmund Koch

## Contents

Instrumentation .....	470
Transversal Resolution .....	472
Axial Resolution .....	472
Wavelength .....	473
FD-OCT .....	473
Light Sources .....	474
Special Considerations of OCT in NDE .....	475
Scanning .....	475
Events .....	476
Applications of OCT in NDE .....	476
Nanoparticles and Polymers .....	476
Functional OCT Imaging .....	478
Biofouling in Water Processing .....	481
Security Applications .....	482
Composite Materials .....	485
Biological Samples and Natural Material Inspection .....	486
Thin-Film Coatings in Various Applications .....	487
Monitoring and Evaluation of Processes .....	492
Optical Surfaces and Samples .....	497
Summary .....	502
References .....	502

---

J. Golde · L. Kirsten · C. Schnabel · J. Walther · E. Koch (✉)  
Clinical Sensing and Monitoring, Anesthesiology and Intensive Care Medicine, Faculty of  
Medicine Carl Gustav Carus, TU Dresden, Dresden, Germany  
e-mail: [Jonas.Golde@tu-dresden.de](mailto:Jonas.Golde@tu-dresden.de); [Lars.Kirsten@tu-dresden.de](mailto:Lars.Kirsten@tu-dresden.de);  
[Christian.Schnabel@tu-dresden.de](mailto:Christian.Schnabel@tu-dresden.de); [Julia.Walther@mailbox.tu-dresden.de](mailto:Julia.Walther@mailbox.tu-dresden.de);  
[Edmund.Koch@tu-dresden.de](mailto:Edmund.Koch@tu-dresden.de)

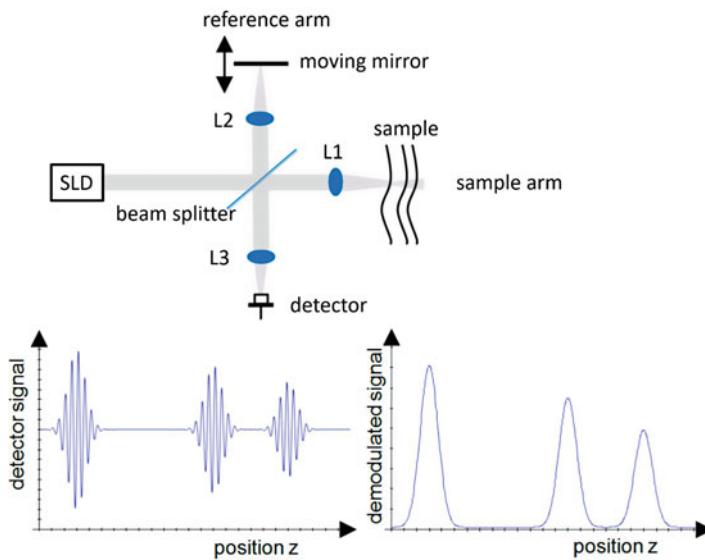
**Abstract**

Optical coherence tomography (OCT) is a noninvasive, high-resolution, interferometric imaging modality using near-infrared light to acquire cross sections and three-dimensional images of the subsurface microstructure of samples. Because of the rapid enhancement of OCT with respect to acquisition speed and axial resolution over the past years, OCT is becoming more and more attractive for applications in nondestructive testing and evaluation. In this chapter, a brief introduction to the technique and the instrumentation is first given, followed by an overview of application of OCT in NDE reported so far.

**Instrumentation**

Optical coherence tomography was developed in 1991 for contactless, noninvasive cross-sectional imaging in biological systems (Huang et al. 1991). The method is based on white light or low-coherence interferometry and was first applied to image the retina and coronary artery. Similar approaches to visualize the retina have been done by Fercher et al. (1988) and Hitzenberger (1991) around the same time. Coherence scanning microscopy, a very similar method used for surface profilometry (Lee and Strand 1990), was introduced at a similar time point and used for integrated circuit inspection. Today, OCT is widely used in ophthalmology and many other medical disciplines. Prominent applications are visualizing human vessels, skin, esophagus, tympanic membrane, lung tissue, and many more. There are several reviews about OCT in biomedical applications (Walther et al. 2011; Shu et al. 2017; de Boer et al. 2017). Compared to the amount of published medical applications, so far, applications in nondestructive testing and nondestructive examination are rare.

Today, the term OCT is used not only for cross-sectional imaging but, similar to ultrasonic imaging, for A-mode, B-mode, M-mode, and volumetric measurements as well. While many types of interferometers can be used for OCT, the most common approach is using a Michelson interferometer using the same beam splitter for dividing and combining the light (see Fig. 1). Within the interferometer, the light of a low-coherence light source, often a superluminescence diode (SLD), is split in a reference and a sample arm. In the sample arm the light is focused on the target, whereby the beam waist limits the transversal resolution. In the reference arm a movable mirror reflects the light, which then interferes in the beam splitter with the light from the sample arm. When changing the length of the reference arm, the intensity in the combined arm oscillates only if light from both arms is within the coherence length of the light source. Thereby, the position of the reference mirror probes the depth of the sample and the intensity of the interference signal is proportional to the amplitude of the backreflection within the sample arm. Plotting the amplitude of the interfering signal as a function of the length of the reference arm yields an A-mode image. Scanning the optical beam over the sample by deflecting the light and decoding the amplitude as the brightness yields a B-scan, which is a



**Fig. 1** Top: Typical setup of a free beam OCT system. The light from an SLD is split by a beam splitter into sample and reference arm. Several layers in the sample contribute to the reflected light, which interferes with the light from the reference arm. As a function of the position of the moving mirror, the detector recognizes interference fringes (lower left) whenever the optical distance between beam splitter and moving mirror matches a reflection from the sample arm. The oscillating signal from the detector is rectified leading to the A-scan signal (lower right). L1–L3 are lenses for collimating the beam

cross section through the sample. Similarly, when the sample or part of the sample is moving relative to the beam the recorded A-scans generate a picture of the motion, which is called M-scan. This scheme can nicely be used to control the thickness and other parameters of a transparent film during production, by placing the OCT optics in front of the transported film (see section “[Nanoparticles and Polymers](#)”). Deflecting the beam in two directions yields volume data that can be sliced in any direction or rendered giving a three-dimensional impression of the sample. A slice through the volume stack perpendicular to the scanning direction is called an *en face* image.

In order to increase the speed of the image acquisition, a larger area of the sample can be illuminated and the single detector exchanged with a one- or two-dimensional camera, allowing parallel acquisition of the interference fringes. This method, called full-field OCT, is more sensitive to sample motion due to the lower scan rate and images may be degraded by multiple scattering due to the loss of the confocal advantage.

Despite faster methods of OCT available today, this method called time domain OCT (TD-OCT) has the advantage that the focus position can be synchronously adjusted with the position of the reference arm allowing optics with a high numerical aperture yielding superior transversal resolution over the full measurement range.

When doing so, one has to bear in mind that white light interference measures the optical length, which is  $n_G \cdot d$ , where  $d$  is the length and  $n_G$  the group refractive index of the material. Moreover, due to refraction the focus position in the material will only vary with  $d/n$ , where this time  $n$  is the (phase) refractive index of the material (Schmitt et al. 1997; Qi et al. 2004).

## Transversal Resolution

Similar to normal microscopy, the transversal resolution ( $\Delta x$ ) of OCT is limited by the numerical aperture (NA) of the sample beam.

$$\Delta x = (2 \lambda / \pi) \text{ NA}$$

The depth of focus, when the focus position is not adjusted to the length of the reference arm, is given by the twofold Raleigh length ( $z_R$ )

$$2z_R = \pi \Delta x^2 / 2\lambda$$

Thus, a high transversal resolution (low  $\Delta x$ ) is associated with a low depth of focus. A typical transversal resolution used in OCT is in the range of 1–20  $\mu\text{m}$ . Due to the limited depth of focus, a typical imaging depth is in the range of 2–3 mm, but ranges around 10 cm have been described (Song et al. 2016; Pfeiffer et al. 2017).

To brake the limit of focus range and transversal resolution given by the Gaussian beam profile, Bessel beams for illumination and/or detection (Lorenser et al. 2014; Yi et al. 2017) and other solutions (Yin et al. 2017) have been used. While maintaining a constant resolution over a larger depth range, a loss of sensitivity is obtained.

## Axial Resolution

Unlike confocal microscopy, in OCT, the axial resolution is decoupled from the axial resolution and depends on the spectral width of the light source. The axial resolution is limited by the coherence length ( $l_c$ ) of the light source, which depends on the spectral width ( $\Delta\lambda$ ) (FWHM) and the center wavelength  $\lambda_0$ .

$$l_c = \frac{2\ln 2}{\pi} \frac{\lambda_0^2}{\Delta\lambda}$$

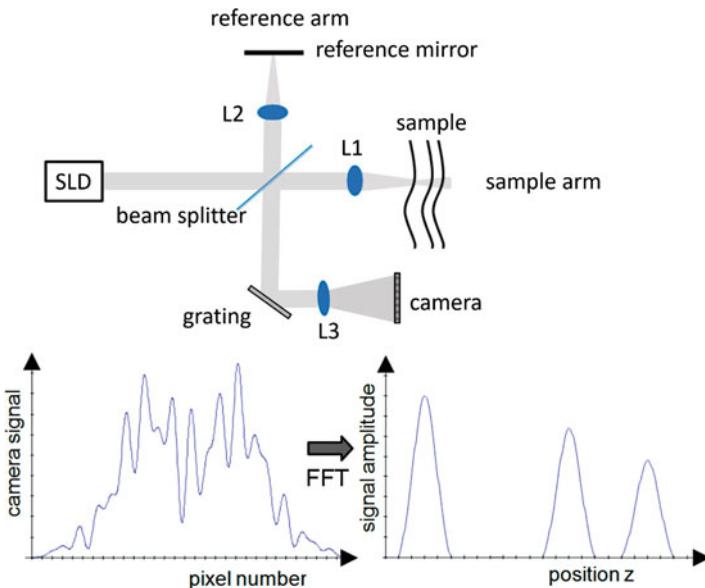
The prefactor depends on the shape of the light source and is minimal for a Gaussian spectrum with a value around 0.44. The large bandwidth of some light sources, in conjunction with the fact that the resolution of OCT in a medium with group refractive index  $n_G$  is  $l_c/n_G$ , allows resolutions in the range of 1  $\mu\text{m}$  and below, while resolutions around 10  $\mu\text{m}$  are more typical (Drexler et al. 1999; Kirsten et al. 2017).

## Wavelength

Due to the higher penetration depth of NIR radiation (700 nm–400 nm) in tissue and the available optical components, two wavelength ranges are mostly used in OCT. Radiation in the 800 nm range allows a higher resolution, while wavelength around 1300 nm enhances the penetration depth due to less scattering in tissue. In NDE, other wavelength ranges, depending on the transmission of the sample, may be beneficial. OCT has been demonstrated from 400 nm (Goldberg et al. 2017) up to nearly 2  $\mu\text{m}$ , which seems to be the optimal wavelength range for many paints (Cheung et al. 2013). OCT with extreme ultraviolet (XUV) radiation with depth resolution of 24 nm will probably have only limited use (Fuchs et al. 2017) due to the few materials transmitting this short wavelength radiation.

## FD-OCT

While the description above explains TD-OCT systems, which have certain advantages, the disadvantage of TD-OCT is the relatively slow speed due to the mechanical scanning of the reference arm. The breakthrough in speed was achieved by introducing the concept of Fourier domain OCT (FD-OCT), which is also known as “Frequency Domain OCT” (Leitgeb et al. 2000) “spectral interferometry” (Fercher et al. 1995), “coherence radar” or “spectral radar” (Bail et al. 1996). There, the interference spectrum is measured with a spectrometer or a tunable light source. The depth-dependent information is recovered by a Fourier transformation of the interference spectrum. Systems with a spectrometer are mostly called spectral domain OCT (SD OCT), while systems containing a tunable light source are termed frequency domain imaging (OFDI). The imaging speed of FD-OCT systems is only limited by the readout rate of the line scan camera of the spectrometer or the tuning velocity of the light source, which is mostly a laser. Due to the additional information about the wavelength of the photons, FD-OCT systems achieve a better signal-to-noise ratio (SNR) in the same time (Leitgeb et al. 2003a; Choma et al. 2003; de Boer et al. 2003). The A-scan rate with spectrometer-based systems (see Fig. 2) can be as high as 200.000/s. While systems containing swept sources have been operated with 20 and recently with 44.5 million A-scans per second (Wieser et al. 2010; Kang et al. 2018), typical repetition frequencies of commercially available swept sources are between 50 and 200 kHz. While the information about the specimen obtained with TD-OCT and FD-OCT is similar, there are certain differences. Because there is no difference between reference and sample arm in FD-OCT, interference between different layers within the sample or the focusing optics will show up in the depth signal as well. There are also differences with respect to axial and lateral movement of the sample, not only between TD-OCT and FD-OCT but even between SD-OCT and OFDI. Due to dispersion in the optics or sample (Fercher et al. 2002; Lippok et al. 2012; Kottig et al. 2012; Pan et al. 2017), uneven spread of the spectrum with respect to the wavenumber  $k$  (Choma et al. 2003; Van der Jeugt et al. 2010), non-Gaussian shape of the light source (Tripathi et al. 2002; Akcay et al. 2003),



**Fig. 2** Schematic of a SD-OCT system. Different than TD-OCT, the mirror in the reference arm is stationary. The interfering light coming from the beam splitter is spectrally resolved by the grating and focused on the line scan camera.  $L1-L3$  are different lenses or objectives. The camera detects the interference spectrum. After a Fourier transformation, the depth-dependent signal is recovered

background noise of the detector, and other issues, the signal processing from the raw spectrum to the depth signal requires several steps to achieve the theoretical resolution and avoid artifacts.

Besides the bulk optics approach for OCT shown in Figs. 1 and 2, systems are often built with fiber optical components simplifying the arrangement of the optics relative to the sample. While this makes the systems much more flexible, care has to be taken to avoid artifacts due to birefringence in the fibers and couplers, as this will influence the polarization in both arms of the interferometer.

## Light Sources

To achieve a high resolution, light sources used for OCT systems should have a large bandwidth and a high lateral coherence at the same time. Due to the relatively low cost and acceptable bandwidth of around 100 nm, superluminescent light emitting diodes (SLD) are preferred. For axial resolution in the  $\mu\text{m}$  range, light sources based on femtosecond lasers (Drexler et al. 1999) and supercontinuum light sources, delivering radiation from 450 nm to more than 2  $\mu\text{m}$ , are suitable, but are significantly more expensive (Szkulmowski et al. 2005; Spöler et al. 2007; Cimalla et al. 2009). Most swept sources are based on semiconductor amplifiers (Choma et al. 2005; Huber et al. 2006; Yasuno et al. 2007).

To achieve higher sensitivity to certain aspects of the sample, many extensions of the principal concept of OCT, explained so far, have been developed. The most essential developments have been Doppler OCT (Wang et al. 1995; Yang et al. 2003; Yun et al. 2004; Vakoc et al. 2005; Walther and Koch 2014; Haindl et al. 2016), being sensitive to axial and with some more sophisticated approaches to all components of the sample movement (see section “[Flow Characterization](#)”) and polarization-sensitive OCT (PS OCT), highlighting the effect of birefringence due to alignment of structures or strain in the sample (see section “[Polarization-Sensitive OCT](#)”).

## Special Considerations of OCT in NDE

Most applications of OCT in biomedical optics have a moderate demand on accuracy as biological data vary considerably. Errors in the scale in all dimensions of about 1% of the full range are mostly acceptable and can be achieved by a simple calibration. In NDE, the requirement on precision may be below  $10^{-3}$  and temperature effects in the spectrometer, aging of (swept source) light sources, scale and linearity of the transversal scanning may become a problem. Especially in high-speed systems, the transversal scanning is often sinusoidal, and the linearization over the full-temperature range may be challenging. Moreover, the deflection of the optics may not be linear and the typical assumption in biomedical optics that the scan is telecentric in spite of the fact that one or both scanner mirrors are not in the focus of the optics might introduce unacceptable distortions. As a typical result of placing the scanner mirrors for x and y on both sides of the focus lens, opposite curvatures for a plane surface in both directions can be noticed. Moreover, alignment errors in the scanning system can lead to tilted coordinate systems (Strupler et al. 2015). Additionally, due to the properties of a Gaussian beam modified by aberrations of the lenses, the resolution may vary over the three-dimensional image area (Sandrian et al. 2012). Both factors will also influence the signal strength, which is also affected by the depth-dependent signal roll off in spectrometer-based OCT and OFDI.

## Scanning

As OCT works in reflection, it requires only one optical port. If the interesting part of the specimen is accessible from one side, telecentric scanning from this side is the preferred choice as all three dimensions are decoupled. This can be achieved by placing a two-dimensional deflecting element in the backside focus of a lens system. Deviations from the focus position lead to field curvature and a fan-shaped pattern, as known from sonography. Fan-shaped scanning patterns allow large fields of view even in the case of restricted optical access as demonstrated by imaging the tympanic membrane through the ear channel (Burkhardt et al. 2014). Moreover, OCT can image a large field of view through a relatively small hole (Burkhardt et al. 2010).

In situations with limited access, rotating beam deflection, as used in cardiology for visualization of stenosis in the coronary arteries, may be useful. The flexible rotating fiber optic catheter may have a diameter well below 1 mm. Three-dimensional imaging is performed by pulling back the catheter, which may be rotating at up to 5000 rotations per second (Wang et al. 2015).

## Events

Although OCT has so far seen only limited applications in NDE/NDT, there have been two conferences OCT4NDT (Linz, Dresden) solely dedicated to this topic. A third meeting planned for 2017 had to be canceled and it is unclear if the series will continue.

The following part of this chapter describes different applications of OCT in NDE/NDT. There might be more applications, which are not published for various reasons. Moreover, there are distance sensors effectively using the principle of OCT, evaluating the frequency of interference fringes, but manufacturers call these systems, e.g., “Spectral Interference Laser Displacement Sensor.” Such sensors are typically used to measure distances or thicknesses only. Some companies build a unit consisting of a light source and a spectrometer, which can be used for confocal chromatic or spectral interference sensors. In this chapter, the focus is on interferometric sensors only. The order of applications described in the following is somehow arbitrary but hopefully representative.

---

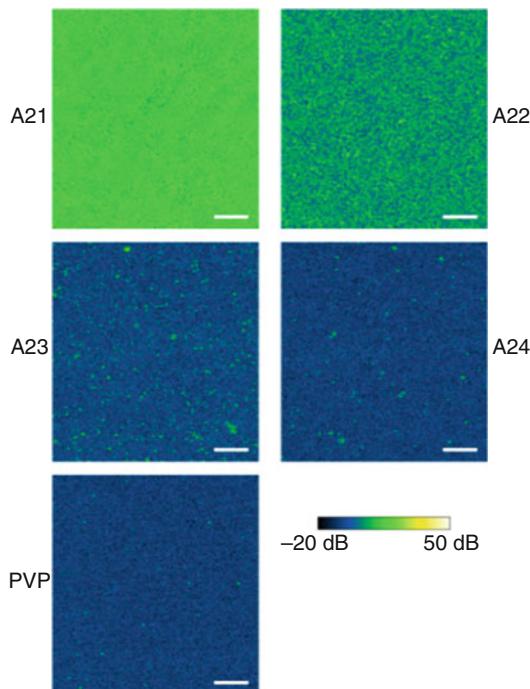
## Applications of OCT in NDE

### Nanoparticles and Polymers

Since nanoparticles can improve material properties, they are widely used as additives for the production of foods, cosmetics, and technical materials, such as nanocomposites. OCT has been reported in exemplary applications as a promising technique for the contact-free and nondestructive characterization of nanoparticles with respect to size, dispersion, and concentration of nanoparticles. A further application is the investigation of the dispersion state by characterizing agglomerates, which are of nonuniform shape and have a size up to several micrometers (Schneider et al. 2016). Thus, OCT is expected to be a potential tool for assessing nanoparticles in the context of quality control and in-line monitoring.

At lower concentrations where the distance between individual nanoparticles is larger than the spatial resolution of OCT, single nanoparticles can be detected with OCT assuming a sufficient backscattering intensity of the nanoparticles. At higher nanoparticle concentrations, a signal enhancement is observed in the OCT images, but single nanoparticles cannot be resolved anymore. Exemplary OCT images for polymer (PVP, polyvinylpyrrolidone) films with different content of gold nanorods are depicted in Fig. 3 (Grombe et al. 2014). The en face OCT images represent areas

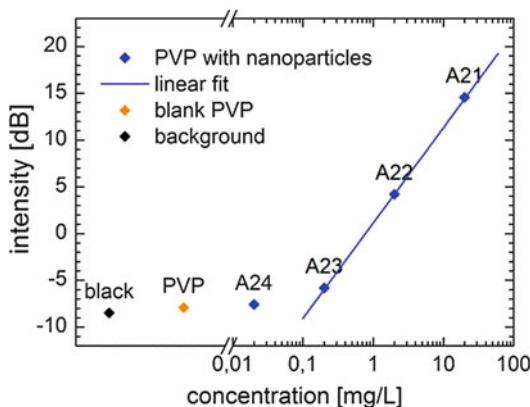
**Fig. 3** (a) The en face OCT images of PVP layers show areas of  $1.2 \times 1.28 \text{ mm}^2$  in a depth of  $15 \mu\text{m}$  beneath the surface for nanoparticle concentrations A21, 20 mg/L; A22, 2 mg/L; A23, 0.2 mg/L; A24, 0.02 mg/L; and PVP, without nanoparticles. Scale bar, 200  $\mu\text{m}$ . (Reprinted from Grombe et al. 2014, Copyright (2014), with permission from Elsevier.)



in a depth of  $15 \mu\text{m}$  beneath the surface of the polymer film. Fig. 4 shows the calibration curve, where the mean intensity is plotted over the nanoparticle concentration. Based on such quantitative characterizations, the determination of unknown nanoparticle concentrations could become possible for specific nanocomposites. The background noise of the OCT system and the background signal of the matrices give a lower detection limit for the nanoparticle concentration, which is 0.2 mg/L for the example depicted in Fig. 4. Microstructures, inhomogeneities, interfaces, and surfaces in the sample cause additional backscattering intensity, which has to be considered since they contribute to the background signal. Furthermore, the calibration curve depends on the specifications of the OCT system, the matrix, the type of nanoparticles, and probably the agglomeration characteristics at different nanoparticle concentrations. Hence, the calibration curves can vary in the bias, the slope, and the detection limit (Grombe et al. 2016).

Polymeric materials and fiber-reinforced composites are further material classes, which are widely used in different application areas due to their improved and tailored mechanical properties. The mechanical and optical properties are influenced by the microstructure and morphology (Hierzenberger et al. 2014). Defects in the sample, such as cavities, micro-cracks, or delaminations, arising from the production process itself or from mechanical loads can degrade the properties and the stability and can reduce the lifetime of materials (Shirazi et al. 2017). The inspection of polymeric samples by OCT allows the assessment of the sample microstructure and the location of defects. For example, OCT allows the measurement of layer

**Fig. 4** Dependency of the measured mean OCT intensity on the nanoparticle concentration for the PVP matrix and formulation used. (Reprinted from Grombe et al. 2014, Copyright (2014), with permission from Elsevier)



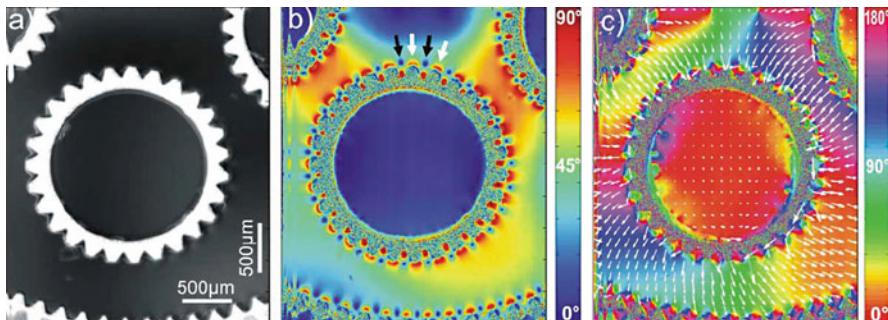
thicknesses in stacked polymeric film samples (Yao et al. 2013) and the investigation of the polymer crystallization process (Hierzenberger et al. 2014). Delamination and micro-cracks cause additional reflections or backscattering, which are visible in OCT images (Stifter et al. 2008; Liu et al. 2014d). Furthermore, the fiber structure and orientation in fiber composites or reinforced materials can be three-dimensionally imaged (Stifter et al. 2008) (see section “Composite Materials”).

## Functional OCT Imaging

OCT is typically utilized for high-resolution structural imaging by means of visualizing the intensity of the light reflected or backscattered from the sample. Additional information about the material properties can be obtained using functional extensions of OCT. This section focuses on the most common techniques established in NDE, which are polarization-sensitive OCT, being utilized for mapping stress and strain in materials, and OCT-based velocimetry for imaging flow characteristics. Some alternative approaches, which are based on speckle pattern analysis (Liu et al. 2014c) or phase-contrast OCT (Zhang et al. 2015), target the imaging of deformations and internal displacements.

### Polarization-Sensitive OCT

Polarization-sensitive OCT (PSOCT) analyzes the polarization of the detected light in order to determine the polarization-changing properties of the sample, mainly birefringence, which occurs, e.g., in cases of internal stress or applied strain. In common PSOCT setups, the sample is probed with circularly polarized light, and the detection unit is extended by a polarization beam splitter in order to measure the interference signals in both, the vertical and horizontal, polarization channels. Processing of both interference spectra reveals two complex-valued A-scans. The amplitude ratio and the phase difference between both A-scans are used to calculate the retardation and the orientation of the optical axis, which both provides information about the stress in the material. The retardation is a measure for the stress in the



**Fig. 5** OCT image of the resist-waver interface of a photoresist mold for a micromechanical wheel. (a) Reflectivity. (b) Color-coded retardation. (c) Color-coded orientation of the slow optical axis and corresponding vector field, scaled by the retardation magnitude. Arrows in (b) show positions at the teeth of the wheel with high strain. (Reprinted from Wiesauer et al. 2006, Copyright (2006), with permission from OSA)

material, which has been demonstrated for plastic materials and polymer samples (Stifter et al. 2003; Wiesauer et al. 2005). However, the retardation measured at a certain position in the sample represents the birefringence along the beam path from the sample surface to the position in the sample. In order to determine the local birefringence, the change of the retardation over depth has to be extracted (Wiesauer et al. 2005; Heise et al. 2010). Additionally, the optical axis orientation indicates the direction of the strain in the sample (Wiesauer et al. 2006).

Thus, PSOCT imaging allows the localization of strained regions in the sample, indicating defects or vulnerable spots. This was demonstrated by Wiesauer et al. (2006) and Heise et al. (2010) for the quality control of molds in photoresist layers, which are used for production of micromechanical gear wheels. There, highly birefringent areas have been located at the teeth of the wheel (Fig. 5). Further applications are the optimized crack detections in glass fiber-reinforced polymer structures (Wiesauer et al. 2007) and the dynamic observation of the internal stress in materials during tensile and fracture tests (Stifter et al. 2010).

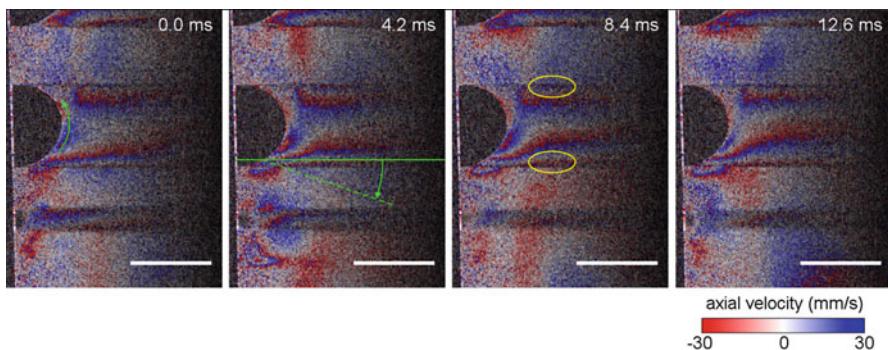
In addition to the retardation and the orientation of the optical axis, PSOCT can provide information about the depolarization, commonly done by calculating the degree of polarization uniformity (DOPU) (Götzinger et al. 2008), which is determined in interrogation windows covering neighboring pixels. A reduced DOPU value is expected in areas with microporosity or micro-cracks due to the accompanying depolarization. Thus, a decreased DOPU can show damages or alterations already before a strong retardation is measured or structural alterations become visible in the PSOCT images, as demonstrated by Stifter et al. (2010) in glass-fiber composite during a fracture test.

## Flow Characterization

The experimental characterization of flows with optical measurement techniques is a widely used approach in experimental fluid dynamics. OCT-based velocimetry is a

proven technique in biomedical imaging and also provides a large potential in the field of NDE. Phase-resolved Doppler-OCT is the most proven approach for measuring motions and fluid flows. If the velocity has a component in the direction of the laser beam, a slight axial displacement occurs within consecutive A-scans. This yields a phase difference between subsequent A-scans. Thus, the velocity component in beam direction (axial velocity) can be calculated from the phase difference between adjacent or subsequent A-scans.

One application is the measurement of small-scale boiling processes at heated surfaces. Boiling phenomena are very effective heat transfer mechanisms with a wide range of applications, e.g., in nuclear reactors or small-scale cooling devices for electronic components. OCT can simultaneously provide the growth behavior of individual vapor bubbles and the fluid flow around the bubbles (Kirsten et al. 2015), which are both important aspects when investigating the microscale heat transfer processes during nucleate boiling. Figure 6 depicts a series of OCT cross sections with a frame rate of 238 Hz showing nucleate boiling in a vertical flow channel on a heated vertical surface. This image sequence has been acquired with a swept-source OCT system providing an A-scan rate of 120 kHz. The experimental setup contains a window, whose inner surface is coated with an indium tin oxide (ITO) layer, which is optically transparent and electrically conductive. Thus, the window can be heated in order to generate nucleate boiling. Simultaneously, OCT imaging is carried out through the window; in Fig. 6, OCT imaging is performed from the left. The ITO layer is visible as a bright vertical line at the left in the OCT images. For visualizing the fluid, an emulsion is used, which contains scattering particles. In contrast, vapor bubbles do not contain scattering particles and are visible as dark semicircular areas at the ITO layer.



**Fig. 6** OCT image sequence showing nucleate wall boiling in a vertical flow channel. The axial (horizontal) velocity component of the fluid velocity is color-coded and was averaged over a  $3 \times 3$  pixel neighborhood before scaling the image. The color scale is repeated for higher velocities. The green arrow (time point 0 ms) indicates the fluid streaming around the bubble. In the second frame (4.2 ms), for one exemplary A-scan (green solid line), the actual optical path (green dashed line) is sketched, which accounts for the refraction at the vapor-fluid interface. The yellow ellipses (time point 8.4 ms) indicate areas, where the highest phase shifts of approximately  $\pm 3\pi$  occurred. (Reprinted by permission from Springer Customer Service Centre GmbH: Springer Nature, Experiments in Fluids, Kirsten et al. 2015, Copyright (2015))

The fluid streaming is visualized in Fig. 6 by color-coding the axial velocity component, which corresponds to the horizontal velocity component in the image. The unambiguous measurement range for the axial velocity is given by a phase difference of  $\pm\pi$ , which corresponds to  $\pm 30$  mm/s for this OCT setup. Red and blue indicate a fluid movement toward the left and right in Fig. 6, respectively. At time point 0.0 s, the green arrow marks the expected fluid streaming around the bubble, where color-coding is in agreement with the direction of the axial velocity component. If the velocity is exceeding the measurement range, a color jump is visible in the OCT image and the color scale is repeated. The highest velocities can be measured behind the bubble, because the laser beam is refracted at the vapor-fluid interface toward or against the streaming direction (upward) in the flow channel changing the Doppler angle (Fig. 6 at time point 4.2 ms). The yellow ellipses at time point 8.4 ms mark the positions, where the highest axial velocities up to  $\pm 90$  mm/s could be measured. In phase-resolved Doppler-OCT, the unambiguous measurement range for the velocity could be further increased using a higher A-scan rate.

Phase-resolved Doppler-OCT is typically used for measuring the axial velocity component; the visualization of two or three velocity components would require multi-beam setups. The visualization of the two velocity components in the OCT image plane is possible using alternative approaches as well, which evaluate the displacement of particles in subsequent B-scans. One possibility is a cross-correlation as used in particle imaging velocimetry, which was utilized by Buchsbaum et al. (2015) for measuring the velocity profiles in a polymer compound during the extrusion process. Furthermore, particle tracking has been demonstrated for visualizing the flow of polymer particles in drying sessile drops in order to investigate the influence on the final drop shape and on the particle agglomeration (Manukyan et al. 2013).

In conclusion, OCT-based velocimetry can be applied in turbid media or in transparent media using tracer particles or contrast agents. OCT allows the spatially resolved measurement of one to three components of the velocity, depending on the technical implementation. In contrast to established techniques such as particle image velocimetry, OCT needs only one optical access, because it detects back-scattered light. This makes OCT a promising measurement technique for challenging experimental setups. Additionally, OCT is well suited for characterizing small liquid volumes and measuring in small geometries.

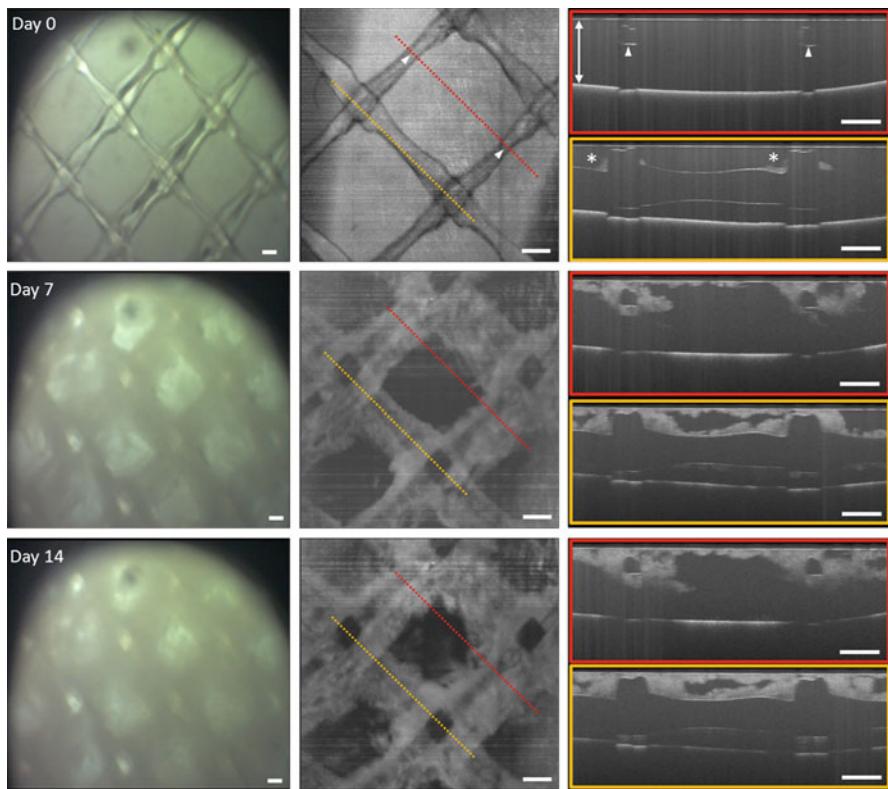
## Biofouling in Water Processing

Filtration systems are of enormous concern in water treatment and drinking water processing. Membrane filtration systems have increasingly been used for high-quality water production in both saline and freshwater environments. A key challenge in all membrane systems is the formation of biofilm on the filtration membrane leading to a decreased filtration performance. This process is called biofouling and is characterized by an increased feed channel drop, permeate flux reduction, and increased salt passage in desalination systems.

To maintain high-efficiency and high-quality water production, fouled membrane elements require periodic cleaning and replacement. For a highly efficient operation of spiral wound nanofiltration and reverse osmosis membrane systems, detailed understanding of biofilm formation and growth is necessary. This understanding is needed on the one hand to optimize the operational aspects of the system regarding, e.g., back-flush or replacement of filtration elements. On the other hand, new developments of membrane and spacer materials, membrane module design, or water treatment before passing the filtration module can only be done with detailed and practical investigations of biofouling formation (Shen et al. 2016; Liu et al. 2017a; Park et al. 2018). Therefore, small-scale and representative test units and in situ nondestructive imaging tools are required. Optical coherence tomography has been successfully used for the visualization of biofilm development and the process of biofouling (Dreszer et al. 2014; Weiss et al. 2016). The advantage is that both 2D and volumetric image information can be repeatedly acquired without any impact on the biofilm and its environment (West et al. 2016; Li et al. 2016; Fortunato and Leiknes 2017; Trinh et al. 2018). Features like thickness, distribution, and reaction on maneuvers like back-flushing can be easily measured and quantitatively compared to gain reliable information about biofouling characteristics (Akhondi et al. 2015). Aside from those obvious advantages, OCT is still limited in penetration depth. Especially in late stages of biofilm development, structures like spacer and membrane are difficult to distinguish. Furthermore, only one layer of the membrane system can be visualized and an in-line application of, e.g., biofilm thickness in industrial filtration systems will be challenging. Nevertheless, OCT allows new insights into biofouling processes and in microorganism growth in small-scale test units and is a valuable tool for the further development of water processing techniques. An example of biofilm formation is shown in Fig. 7. The microscopic top views in the first row of Fig. 7, acquired with a video camera, show the spacer network with the filtration membrane beneath. The OCT cross sections in the second row show the advantage and limitation of this imaging technique. Without any preparation, OCT allows the visualization of the feeding channel, the spacer, and the filtration membrane. The backside of the spacer cannot be visualized due to scattering loss of the penetrating OCT light. Moreover, in the presence of a massive biofilm, the spacer structures become hardly distinguishable from each other, and further image processing is necessary for quantitative measurements like microorganism volume and distribution.

## Security Applications

Another emerging field for OCT is the security and defense sector, where the need for reliable detection systems became more and more important over the last decades. The main applications can be found on the one hand in biometric identification modalities and on the other hand in the detection of security features in valuable documents like passports or credit cards and keycards, in paper money and banknotes and in video and image protection.



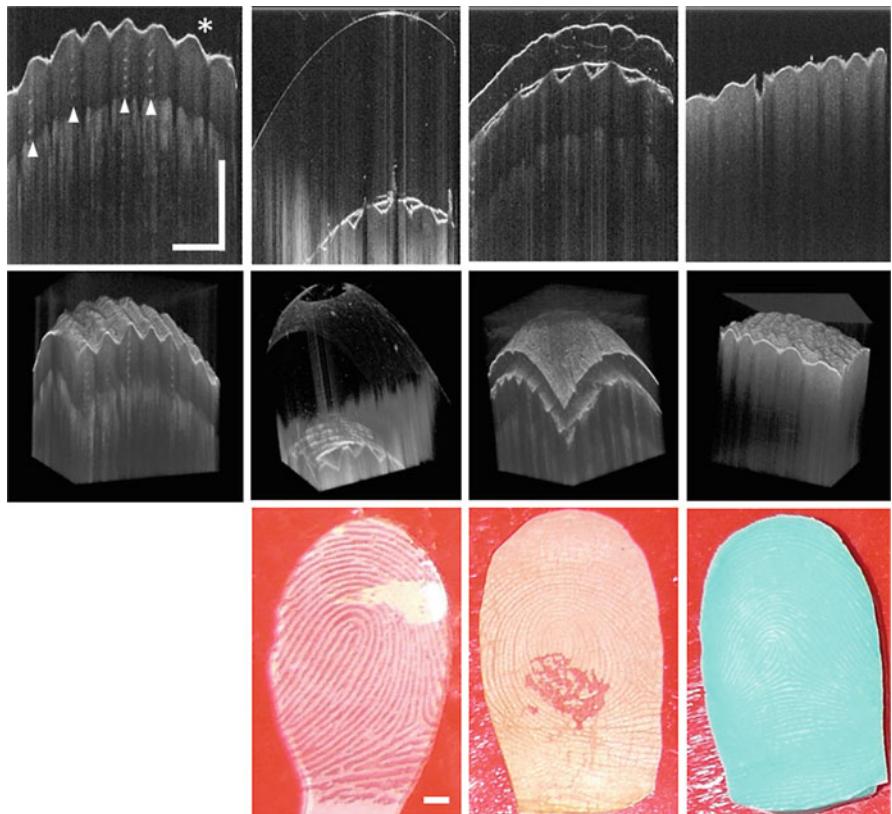
**Fig. 7** Imaging of a small-scale test unit with a filtration membrane and the spacer mesh in the feed channel (approx. 600 µm marked with the arrow). Video microscopy (first column) and OCT en face (second column) and cross sections (third column) show the biofilm distribution and thickness. From top to bottom: no biofilm (day zero; fibers marked with triangle and sealing artifacts at fiber mesh nodes marked with asterisk), medium biofilm after 7 days and biofilm after 14 days. Due to different refractive indices, it appears with an axial shift beneath the spacer. Scale bar corresponds to 400 µm

Fingerprint identification has been adopted all over the world by government agencies (e.g., airport, police) and is also part of our private environment (e.g., smartphones, laptops). The database of fingerprints is the most comprehensive biometric database for personnel identification and this technology plays a key role in biometric-based security applications. However, wet, dry, or damaged fingertips may result in bad image quality, and current fingerprint acquisition systems, often based on two-dimensional photo images, lack robustness in those cases. Moreover, they are vulnerable to identity theft and spoof attacks using fake fingers. The utilization of thin-layer fingerprint fakes is a well-established method to overcome conventional commercial fingerprint scanners based on 2D image acquisition. These thin-layer fakes can be produced very easily in a short time and at a low price.

OCT is a promising tool to overcome those drawbacks of current technologies. It allows for contactless acquisition of three-dimensional high-resolution images of the

internal fingerprint structures with all the features of a human finger (finger pulp geometry, multilayer texture, and respiratory ducts channels) (Meissner et al. 2013) and additionally viable skin tissue parameters like blood flow in the microvasculature (Liu et al. 2017b). Hence, OCT is a very robust technology with respect to both different image quality conditions and potential spoof attacks. A comparison of different thin-film fingerprint fakes and the physiological human fingerprint is shown in Fig. 8 as an example. The human skin consists of three layers: the epidermis, the dermis, and the subcutis. The dermis contains the respiratory and sebaceous glands connected by the respiratory ducts to the pulps of the epidermis surface. The substrates of thin-film fakes are homogeneous and unlayered and no spiral ducts are visible.

The technical effort and financial cost associated with 3D OCT imaging is one of the drawbacks of this technique, which hampered the breakthrough in biometrics and

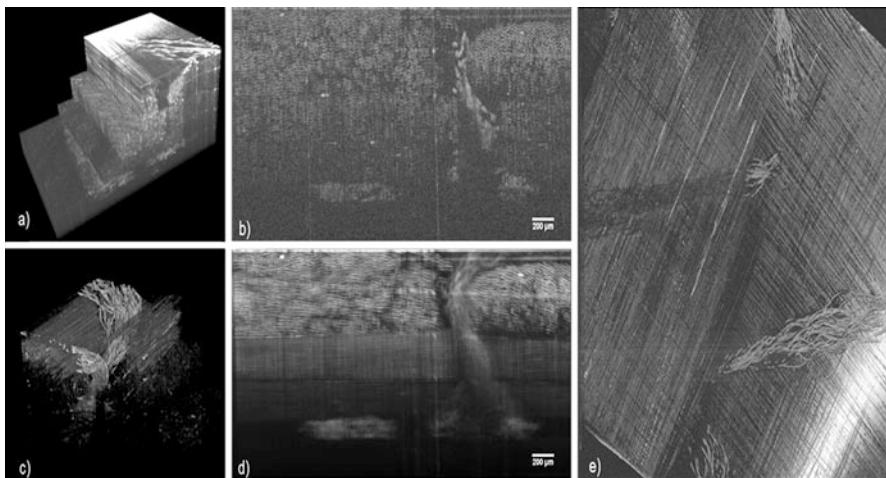


**Fig. 8** Comparison of real human fingertip and thin-layer fakes. OCT cross sections (first row) and 3D representation clearly shows the advantage of high-resolution OCT imaging of the skin microstructure (fingertip pulps marked with asterisk and respiratory ducts marked with triangle). The fake fingerprints can be clearly distinguished from real human skin. Scale bar corresponds to 1 mm

other security applications. Nevertheless, there are other nonscanning cost-efficient adaptations of OCT in security applications. In electronic card readers, the user inserts or swipes the card, and this movement can be used in a one-dimensional common-path OCT system to acquire a two-dimensional cross section where different security features like layer encrypted codes can be extracted. This can also be achieved by measurements of A-scans in which a different layer thickness is used for the encryption of the valuable information and the interlayer border is acquired with OCT (Gandhi et al. 2015). Two-dimensional OCT cross sections were also successfully used to monitor security features in paper money (Choi et al. 2010; Fujiwara and Matoba 2011a, b).

## Composite Materials

The use of composite materials is of increasing interest due to the rising requirements of efficiency and reliability in modern high-performance rotating blades like those used in gas and steam turbines, wind power plants, and pumps. Because conventional materials are not suitable to address these increasing requirements, newly developed composite materials can overcome previous restrictions. The tunable successive damage behavior of multilayer materials is especially advantageous for the design of rotor blades with a docile and announcing malfunction characteristic. Structural integrity is hampered either by the long-term strain during life cycle or by additional impacts, e.g., animals (Liu et al. 2014b). Therefore, different damages in the layer setting occur and influence the fiber mesh and linking materials. Those damages can be porosity, cracks, debonding, fiber breakage, and delamination. A detailed knowledge about damage formation and structural disorders is necessary to optimize components life cycle and to achieve cost-efficient and secure turbine operation. Because vibration characteristics change with different and increasing structural damage, the beginning of those changes needs to be visualized. While ultrasonic scan and x-ray computed tomography are mainstream techniques in nondestructive testing of composite materials, OCT has been established for the characterization of fiber structures and damage characterization in ex-situ experiments (Liu et al. 2014a; Dunkers et al. 1999). Ultrasonic scans cannot provide high-resolution image quality and usually require some coupling agent. X-ray is cost-intensive and not adequate for inline inspections. OCT as a high-resolution nondestructive imaging technique is a promising tool for the visualization of different structural failures like crack and crack growth and delamination processes during static and even, with further technology development, in dynamic high-performance experiments (Stifter et al. 2008). As an example of high-resolution OCT imaging, Fig. 9 depicts a cutout of a rotor blade made from fiber composite material. The different orientations of the layers become clearly visible; also the course of the supporting fibers can be seen. Due to the volumetric OCT scan, the material formation can be analyzed in detail, and different views can be generated in the offline data processing for the measurement of layer thickness, fiber length, fiber cracks, and delamination size.



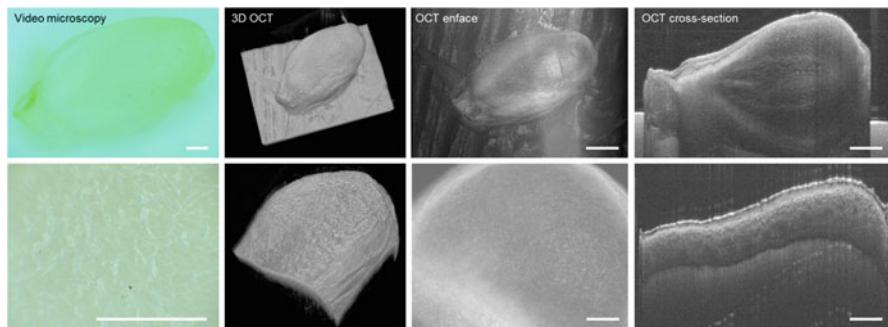
**Fig. 9** Cutout of a rotor blade made from fiber composite material visualized with OCT. (a) to (c) show the advantage of high-resolution 3D imaging. Different views can be generated from the 3D OCT data cube like the en face view (e) or the summarized image of 250 cross sections (d). Imaged area was  $3 \times 2 \times 1.4 \text{ mm}^3$

## Biological Samples and Natural Material Inspection

In addition to the wide range of biomedical applications, OCT is increasingly used for the characterization and monitoring of biological samples and natural materials.

In agriculture, high-quality seeds are one of the most valuable items of trade and therefore, quality control, either during seed processing or for further development regarding germination or environmental resistance, is of great interest (Fan and Yao 2012; Wijesinghe et al. 2017a, b). Furthermore, large amounts of seeds are stored for long time periods in silos and an early detection of potential diseases and infections (like viral infections) is necessary to prevent losing the whole harvest (Lee et al. 2011, 2012). OCT provides an easy way to measure structural features like the gap distance between seed coat and endosperm or the thickness of various morphological layers (Verboven et al. 2013) (refer to Fig. 10). Therefore, measurements can be done repeatedly to study the influence of different chemical treatment procedures on seed germination (Wijesinghe et al. 2017b). Also quality control of vulnerable fruits like kiwifruit, mandarins, and loquats was successfully achieved using OCT (Magwaza et al. 2013; Li et al. 2015; Zhou et al. 2018). In those applications, contactless and fast imaging techniques are especially necessary. OCT allows such measurement and analysis of individual fruits, reduces waste, and permits repeated measurements without any preparation. The structural depth-resolved information from OCT cross sections in combination with computational analysis models allows fast and reliable characterization of fruit skin and enables OCT for inline quality monitoring in the agriculture industry.

Another application of OCT for structural inspection and distance measurements comes from the poultry industry, where the egg shell thickness and shell crack



**Fig. 10** Video microscopy and OCT representation of cucumber (first row) and melon seed (second row). Scale bar corresponds to 500  $\mu\text{m}$

formation are important quality parameters to reduce problems like breaking, microbial contamination, and water vapor loss during incubation (Sabuncu and Akdoğan 2014, 2015).

Quality control and material characterization also play an important role in processing industry and handcraft applications, either for studying material characteristics and behavior on environmental impacts (e.g., paper quality inspections using 2D and 3D OCT imaging (Kirillin et al. 2007, 2008; Prykäri et al. 2010); characterization of porous materials like sandstone (Bemand and Liang 2013; Campello et al. 2014)), for the evaluation of material appearance after processing (protective wood coatings (Latour et al. 2009; Dingemans et al. 2015); or for valuable arts like jade figures (Chang et al. 2010; Sun et al. 2016). Therefore, the structural information (e.g., layer thickness) in combination with the analysis of optical properties like reflectivity, which changes due to the composition of protecting coatings or in the presence of substances, which influence the scattering behavior like water in porous sandstone, lead to a noble way of nondestructive evaluation and analysis using OCT.

## Thin-Film Coatings in Various Applications

### Tablets

OCT as a nondestructive imaging modality has recently been used for quality control in the pharmaceutical industry. There, the process of coating one or more layers of polymer onto tablets is one of the main issues in pharmaceutical manufacturing in order to achieve uniformity in color, light protection, and taste masking but also advanced coatings such as active coatings and sustained release to increase the therapeutic efficacy of tablets (Felton 2016) are addressed. In general, methods for manufacturing homogenous coatings on a tablet are spraying a liquid, dipping into liquid, or depositing powder by electrostatic forces whereby spray coating is the most widely used method to apply polymer film of various compositions onto tablets. Although coating techniques are well established, there are unresolved issues

of influencing factors causing difficulties such as nonintact, cracked or roughened coatings, tacking tablets, coating inhomogeneities and film thickness variations and these affect the final tablet quality. Therefore, average coating thickness and intra- and inter-tablet coating uniformity are the issues of high interest to be quantified. By definition, inter-tablet coating uniformity corresponds to the coating thickness variations between different tablets within a batch, while the intra-tablet uniformity refers to the coating thickness variation on a single tablet. To ensure a high and consistent quality across the batch, inter-tablet variations should be low. In addition, intra-tablet uniformity is highly important for functional film coatings with sustained release where the coating thickness is mainly responsible for the drug release rate.

For the quantification of intra-tablet coating uniformity, various measurement techniques are applied such as laser-induced breakdown spectroscopy or nondestructive modalities like hyperspectral near-infrared imaging, ultraviolet chemical imaging, X-ray micro-computed tomography, terahertz pulsed imaging (TPI), and, more recently, OCT. OCT has the advantage of providing a high-data acquisition rate and simultaneously a high spatial resolution. For these reasons, it seems to be a suitable modality for the nondestructive evaluation of pharmaceutical coatings. To date, only a few coating formulations have been measured by OCT. Lin et al. have examined 22 different common coatings with different pigments and formulations to evaluate OCT as a nondestructive method for thin-film coating measurement (Lin et al. 2017b). Not surprisingly, the coating transparency for OCT is mainly dependent on the pigment type, coating formulation, and tablet core as well as the coating process itself. Especially for the case of included titanium dioxide, the coating thickness measurement is not possible due to a blurred coating core interface. On the other hand, for coatings without titanium dioxide, the coating transparency is not always given. The authors summarized that OCT is useful for evaluating functional polymer coatings without any pigmentation in the coating itself but also different approaches of the manufacturing process. For the case of tablets with pigment, TPI is recommended as an alternative method for coating thickness greater than 30 µm.

With regard to the coating process, Koller et al. (2011) have characterized pharmaceutical tablet coatings by imaging them at different stages of an industrial drum spray coating process ranging from uncoated tablets up to tablets with a coating thickness up to 70 µm. Similar to other basic research experiments for OCT tablet thickness determination (Dong et al. 2017), 3D imaging of the tablet surface was achieved by a motorized translation stage for moving only a very small number of tablets orthogonal to the scanning cross section. As a result of the polymer coating imaging at different stages of the coating process, OCT is highly suitable for imaging thin layers of less than 20 µm and of course the increase in coating thickness. Besides the processing control, inter-tablet uniformity is exemplarily presented by means of five tablets to estimate the reliability of the coating process with the result of a large variation in the coating thickness from tablet-to-tablet and with this a low inter-tablet coating uniformity. Remarkably, OCT also enables the imaging of inhomogeneities caused by air inclusions in the coating or surface roughness of the underlying tablet core material.

To clarify the potential of OCT as an in-line quality control tool, Markl et al. (2014) have examined the influence of the image quality of the tablet coating with regard to the significant signal power decrease due to a dominant transverse motion of tablets relative to the OCT scanner head occurring during in-line monitoring with SD OCT. Predictably, the determination of homogeneity of the coating turned out to be difficult with increasing transverse velocity of the tablets while the measurement of the coating thickness was still highly accurate. Consequently, for in-line quality control of tablet coatings, the scanning speed and scanner head design have to be adapted. Besides the fundamental experimental studies of the coating process on single tablets, in-line monitoring of randomly moving tablets and pellets in fluid bed coater (Markl et al. 2015b) during the coating procedure has already been realized in laboratory scale (Lin et al. 2017a) since a large number of measured tablets can provide substantially more information about the inter-tablet uniformity. Despite the broad thickness distribution found in this study, the experiments prospectively increase the understanding of the coating process itself with the aim to achieve a uniform coating distribution throughout the batch.

Basically, another challenge is the data acquisition and processing of the OCT datasets of the highly resolved tablet surface for the estimation of the intra-tablet uniformity. Usually the measurement of the intra-tablet uniformity by OCT is determined by means of a limited number of B-scans, which unfortunately represents only a fragment of the entire tablet surface (Koller et al. 2011; Lin et al. 2015). For this reason, Dong et al. (2017) have developed an algebraic reconstruction technique (ART) for OCT imaging of intra-tablet coating uniformity with an increased data management while achieving a significant reduction of the required data without affecting the axial resolution and imaging depth. ART (Leitgeb et al. 2003b; Seck et al. 2011) is based on a similar concept to the recently developed compressive sensing technology (Liu and Kang 2010; Liu et al. 2012), which allows the A-scan reconstruction from raw randomly sampled interferograms without the need of data interpolation. By using an orthogonal matching pursuit (OMP) in combination with the spectrum split approach, the reconstruction is speed up. As a result, the comparison of the OMP algorithm using only 30%, 50%, and 75% of the randomly selected measurement data has shown that the usage of 30% of the dataset for image reconstruction leads to no significant loss of information in terms of tablet coating thickness. Moreover, 30% of the dataset is necessary in order to reconstruct the depth profiles of the whole tablet.

In summary, the accurate measurement of the coating thickness is possible by SD- and SS-OCT as long as an appropriate contrast between the junction of tablet coating and tablet core is achieved. Since coating quality, e.g., coating thickness and homogeneity, can be investigated independently of variations of the tablet bulk material, OCT turned out to be a promising tool on the basis of fundamental research studies. It was also shown that OCT has a much smaller measurement error in comparison with indirect standard methods like weight measurement and diameter determination, where no discrimination between tablet coating and core is possible. Last but not least, OCT can be applied as a low-cost instrumentation for coating quality evaluation during and after manufacturing. In terms of future automated

quality control, numerical simulations are recommended to be performed and compared to the experimental data as well as further developments for faster in-line applicable systems (Markl et al. 2015b) in combination with automated algorithms (Markl et al. 2015a) have to be realized.

## Pearls

OCT has also been used to evaluate the quality of pearls including nacre and nucleus in a nondestructive way. Currently, huge yields of freshwater pearls are produced for which reason the pearl quality inspection has received more attention. In general, the structure and the characteristics of the nacreous layer are crucial factors for pearl grading. Moreover, the nacreous layer is thicker in beadless pearls in comparison with beaded ones. The essential component of the pearl nacre is calcium carbonate whose polymorph form aragonite is the main form occurring stable phase. Unfortunately, during the pearl growth, calcite impurities including traces of organic matter could be formed in the nacre layer of beadless pearls, which results in a decrease of the pearl quality. Consequently, a nondestructive and efficient method for the internal quality evaluation of pearls, including shape, luster, and surface defects, is of high interest but still difficult (Zhou et al. 2016). The typical way to perform the grading of pearls is realized by a gemologist who evaluates the shape (balance), luster, and thickness of nacre, while the latter is the most important one. Currently, X-ray radiography and electron microscopy are used for determining the appearance of the nucleus and the laminated pattern of the nacre (Wehrmeister et al. 2008). Unfortunately, pearl quality measurement by X-ray radiography is limited due to its projectional characteristic. In comparison, electron microscopy has the drawback of being a destructive modality since the pearls have to be cut off and consequently lose their value. As a nondestructive imaging modality, OCT seems to be promising for, e.g., classification between beaded and nonbeaded pearls, detection of internal pearl defects, identification of the pearl nuclei, and the measurement of the nacre thickness (Zeng et al. 2008; Lei et al. 2009; Ju et al. 2010). In general, OCT has the advantage of being a nondestructive, high-speed, high-resolution imaging modality, which can be easily applied to almost all kinds of pearls. First fundamental experiments using OCT for grading the pearl were performed by Ju et al. (2010). There it was shown that OCT enables the differentiation between beaded and beadless pearls in addition to thickness measurement and substructure imaging of the nacre layer. Moreover, internal defects such as cracks in the nacre and nucleus of the pearl were visualized three-dimensionally. In terms of pearl production and monitoring by OCT in practice, Liu et al. (2013) have proposed an effective method for automated measurement of pearl thickness on the basis of OCT cross-sectional images. There, a total variational (TV) model filter has been applied for image denoising followed by the usage of a self-adapting Canny operator for edge detection. Last but not least, circle fitting technologies are used for detecting the boundaries and finally calculating automatically the nacre thickness. Beyond that, the group around Zhou et al. (2016) has used OCT for the inspection of internal defects of beadless Chinese ZhuJi freshwater pearls in combination with a self-developed fully automated algorithm to differentiate between normal and defective sublayers of calcite in the nacre.

There, the proposed algorithm is based on the graph segmentation approach relying on the vertical gradient changes to determine the boundaries of defect sublayers to the neighbored normal layers. The results have shown that the developed approach achieved effective distinction between structural imperfection in the nacre layer of beadless pearls with high sensitivity and specificity.

### **Coatings/Paints**

OCT has also been successfully applied for the noninvasive imaging of subsurface microstructure of painted art objects as well as the monitoring of historical artworks and monitoring of conservation processes. Specifically, the suitability of the proposed method for different art applications was shown by imaging of underdrawings (Adler et al. 2007) and paints with varnish layers (Targowski et al. 2006b; Spring et al. 2008) but also dynamic monitoring of laser ablation of varnish layers (Gora et al. 2006) as well as observing of canvas deformation caused by humidity changes (Targowski et al. 2006a). Since one limitation of OCT is the low penetration depth due to scattering and absorption, paintings are preferentially imaged nondestructively by OCT at a center wavelength of 2.2  $\mu\text{m}$  due to the low water content in comparison with biomedical samples (Liang et al. 2013). On the basis of that, Liang et al. (2015) have developed an FD-OCT system using a supercontinuum laser source with a bandwidth of 220 nm and a center wavelength of 1960 nm. In the corresponding study, different paint samples consisting of historical artist's pigments such as cobalt blue, titanium white, yellow ochre, and indigo in oil painted on a glass microscope slide have been examined. As a result, efficient 3D imaging at a high axial resolution of 6  $\mu\text{m}$  in paint was shown since the paints are transparent at 1960 nm. Furthermore, the contrast of OCT images and with this the layer segmentation suffers from speckle noise. For this reason, Hughes et al. (2010) have proposed a method for speckle reduction based on angular compounding with the result of an improved speckle contrast ratio in cross-sectional images. Consequently, thin layers become more distinct and edge detection algorithms become more efficient. Summarizing the above, the combination of a varying angle of incidence of the sample beam in combination with an OCT system centered at 2  $\mu\text{m}$  seems to be promising for prospective applications on historical paintings with typical paint layers of interest up to 100–200  $\mu\text{m}$ .

### **Protective Coatings/Corrosion Inspection**

Most widely used techniques for the evaluation of anti-corrosion protective coatings are invasive and usually need special preparation of the tested sample. One of the most serious challenges is the detection of any degradation in coating thickness and quality caused by pitting corrosion. Therefore, OCT is a useful modality for the inspection of painted anti-corrosion protective coatings and corrosion progress monitoring. Moreover, surface defects and pits but also internal structures can be imaged (Antoniuk et al. 2012; Rodriguez et al. 2011). There, different anti-corrosion coatings including acrylic paint, Hammerite paint, and multilayer lacquer as well as polymer coatings were evaluated, whereas OCT is mostly promising for paintings in the transparent variant or less pigmented. As a selected example in detail, long-term

conservation of museum objects such as old engines and machines for textile production, mining, or agriculture, which represents important technological developments of the industrial era, is realized by transparent protective coatings of sufficient thickness, which are necessary to maintain the historical character and save the cultural heritage. Since the early nondestructive detection of coating impurities is most important to prevent further damage, OCT might be a promising modality for determining the coating thickness after the application as well as for follow-up monitoring over time and for differentiating between corroded and non-corroded areas (Lenz et al. 2017). Commonly, electrical impedance spectroscopy (EIS) is used for estimating the functionality of the protective coating with layer thickness on the atomic scale. Since only qualitative coating thickness is provided in conjunction with time-consuming and low-resolution measurements, possible improvements could be enabled by OCT. In the study of Lenz et al. (2017), OCT has been evaluated in comparison with magnetic inductive (MI) method by means of customized samples comprising bare metal and polymer coating layers of varying thicknesses. Since MI does not allow the coating thickness measurement above a corrosion layer due to its high inaccuracy, OCT is applied solely for detection of the coating thickness with underlying corrosion. From the results, the potential of OCT as a prospective device for the evaluation of transparent coatings and the differentiation of the underlying metal and corroded material is convincingly shown by means of the high-accuracy and high-speed detection in comparison with the commonly used MI thickness measurement. As follow-up to the fundamental measurements, long-term monitoring of coated metal samples is advisable to evaluate the coating quality in terms of performance and long-term stability.

## Monitoring and Evaluation of Processes

Since its introduction, the acquisition speed of OCT has evolved by a factor of several hundreds, now offering the recording, visualization, and analysis of entire volumes at video rates. This has opened the field of in-line process monitoring toward the application of this method in real manufacturing processes. While most published applications still have been demonstrated in a static or simulated setup, the further development, aiming for the implementation in a dynamic environment, is outlined in the majority of reports.

Besides several highlights of process monitoring that are discussed more in-depth subsequently, some applications will be mentioned, which represent the broad variety of possible applications of OCT in this field. The potential of OCT, for instance, has been demonstrated for monitoring the progress of cure in an adhesive, where it enables the spatially resolved assessment of local imperfections and the relative viscosity in different sample regions via a time-dependent cross-correlation analysis (Ford and Tatam 2013). Similar results have been obtained in a study on crystallization processes in polymers (Hierzenberger et al. 2014) that applied polarization-sensitive OCT to directly monitor the structure development throughout the whole experiment from melt to final cooling with additional information on the

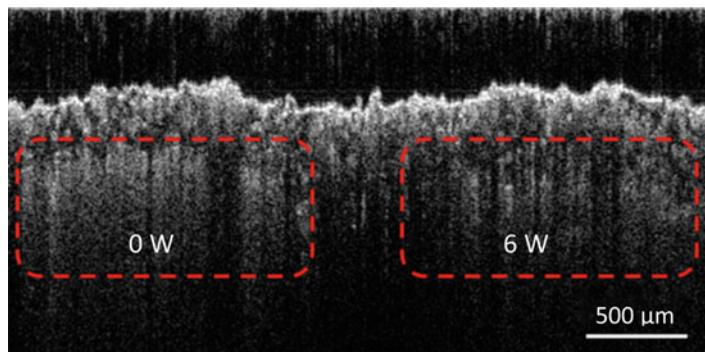
birefringence of the pertinent structures. The identification and automatic quality inspection of weave patterns were presented in another study that utilized OCT with a customized evaluation algorithm for the probing of checked and colored fabrics (Sabuncu et al. 2017).

Furthermore, OCT has been reported in several cases for the analysis of microfluidic devices, where it offers not only velocity profiles by the phase-resolved Doppler evaluation, but can also assess the performance of particle sorting (Dong et al. 2016). In the latter case, OCT enables the real-time functional analysis of an inertial microfluidic system due to its simultaneous high-speed and high-resolution, label-free imaging.

### Material Machining and Processing

In the broad field of material processing, there are several applications that make use of the capability of OCT, to measure both surface and subsurface structures in the micrometer range at a high speed and therefore, to enable the real-time assessment of process parameters and defects. Additive manufacturing is an emerging technique that provides a fast and cost-efficient fabrication of prototypes as an alternative to traditional, subtractive manufacturing. Currently, the field is transitioning toward scalable and high-value manufacture, which requires in-process monitoring and closed-loop machine operation to achieve reliable quality of final products. In case of selective laser sintering (SLS), a technique that utilizes a high-power laser to sinter a powdered material (typically plastic, metal, ceramic, or glass) in the form of a predetermined 3D model, OCT has been demonstrated as a valuable tool to detect loose powder in a solid specimen (Guan et al. 2016). As it is exemplarily presented in Fig. 11 for powder areas with and without laser melting, OCT can thus enable an optimization of the applied power. Current limitations and ongoing research include the improvement of imaging depth and the understanding of the scattering behavior of the utilized powder particles that exhibit an inter- and intra-particle multiple-scattering, resulting in imaging artifacts and reduced penetration depths (Lewis et al. 2018). An advantage of OCT is thereby the possibility, to directly integrate OCT into the sintering beam path, which has recently been demonstrated in a running system (Gardner et al. 2018) and is realized by a dichroic mirror that separates spectral ranges of probing OCT beam and sintering high-power beam. A synchronized and real-time monitoring can thus be ensured and allows an in-line control of laser power and heating parameters.

With comparable approaches, OCT has been integrated in the field of laser structuring and welding, providing the same key aspects of surface and depth-resolved monitoring as in sintering processes. Besides combining beam paths, it was shown that imaging and micromachining with the same light source is feasible, if the broadened spectral part of a short pulse optical source, typically undesirable in a machining laser, is accordingly utilized (Webster et al. 2007). Ablation dynamics were observed in an OCT M-mode, where scanning mirrors are stationary during the machining of a certain position and the OCT measurements provide a depth- and time-resolved analysis of the progression. The spatial and temporal synchronization of machining and probing is therefore an intrinsic property of this approach.



**Fig. 11** OCT cross-sectional structural image of areas with different loose powder density caused by shifting laser energy: (left) area with no laser melting, (right) area with 6 W laser. (Published as Fig. 7c in Guan et al. 2016 under the CC-BY-4.0 license [<https://creativecommons.org/licenses/by/4.0/>])

As one requirement for the integration of optical monitoring technologies into the machining optics is the optics optimization for both utilized spectral ranges, the combination of processing and probing in the 1000 nm range seems feasible. Several publications toward integrated process metrology in laser surface structuring and laser transmission welding are based on this optical window, as a study on a combined laser structuring and Fourier domain OCT system demonstrated, which utilized a nanosecond pulsed fiber laser with the central wavelength at 1064 nm for processing and a superluminescent diode (SLD) with a spectrum of  $1017 \pm 50$  nm for monitoring (Schmitt et al. 2012). Due to significant differences between the incident power of the processing and the monitoring light source, here 20 W in contrast to 10 mW of the SLD, attention has to be paid to the coupling concept in order to prevent excessive power in the imaging system, which may cause measurement artifacts or can damage optic and detection components. The use of an optical edge filter (Schmitt et al. 2012) accordingly allows the adjustment of the edge frequency to separate the processing laser in transmission from the measurement beam, which is reflected into the OCT setup. Additional filtering can improve the achieved extinction ratio.

The same group, moreover, has demonstrated that this approach is also applicable in other laser processing techniques, as, for instance in laser transmission welding, where a laser at 980 nm is utilized for processing and a swept source laser around 1060 nm allows high-speed monitoring of the weld seam and quality control of polymer welding processes (Schmitt and Ackermann 2016). The measurement thereby helps identifying occurring defects and holes during the running process as well as pre- and post-control without additional part handling.

In silicon wafer sawing and polishing, low coherence interferometry (LCI) as an OCT-related technology is widely used for high-accuracy surface topography measurements. For the monitoring of processes during wafer sawing, LCI was demonstrated in the near-infrared wavelength region, which allows surface inspection from

the inside due to the sample transmission in the utilized range (Gastinger et al. 2011). Chemical mechanical polishing, on the other hand, is a key process for the global planarization of silicon wafers, where a wet pad of particular roughness is rotated and wetted with a slurry feed, while silicon is pushed against the rough surface for polishing. As the performance of the process is mainly influenced by the surface texture of the wet pad, in-line information about the pad dynamics are of great interest. It was demonstrated that full-field OCT, an expansion of TD-OCT for scan-free en face imaging with high lateral resolution, is suitable in providing this information *in situ* for a more efficient and productive polishing process (Choi et al. 2011).

### LEDs, Solar Cells, and Displays

The nondestructive, three-dimensional evaluation of layered, electronic components is still an emerging field, as the most well-established probing technologies offer a high-resolution, in-depth analysis only by destroying the examined sample. Several concepts for the OCT evaluation of light-emitting diodes (LEDs), solar cells, and displays have been proposed so far in a post-processing environment but exhibit the potential to also be applied under in-line conditions.

In the case of chip-type LEDs, wire connections can be monitored by OCT in the 1300 nm range, whereas the distribution of fluorescence material is more clearly observable around 850 nm (Cho et al. 2012). As presented in numerous studies, the integration of GPU processing in the OCT workflow hereby enabled the real-time displaying of three-dimensional datasets, aiming for an improvement of the inspection efficacy over traditional inspection methods such as CCD camera or X-ray instruments.

A similar approach is also applicable to roll-to-roll coated polymer solar cells with a multilayer structure, which is prepared by sequentially applying each layer in discrete coating and printing steps (Thrane et al. 2012). Occurring coating and printing errors thereby lead to defects that may disrupt solar cell function or simply cause a poor esthetic appearance of the device. Although the authors mentioned current limitations due to the resolution in the range of microns, they were able to provide three-dimensional images of an encapsulated solar cell module, comprising barrier layers, adhesive layers, substrate, and electrodes.

Additional to the 3D microstructure, a phase-sensitive evaluation of the OCT data was utilized by another group to analyze the nanostructure of certain layers with an inverted pyramid structure, which is commonly used for the reduction of interface reflections (Tsai et al. 2015). Furthermore, the implementation of a second reference arm allowed the quantitative determination of reflection coefficients, which are connected to the nanostructure as an important parameter for the efficiency of solar cells.

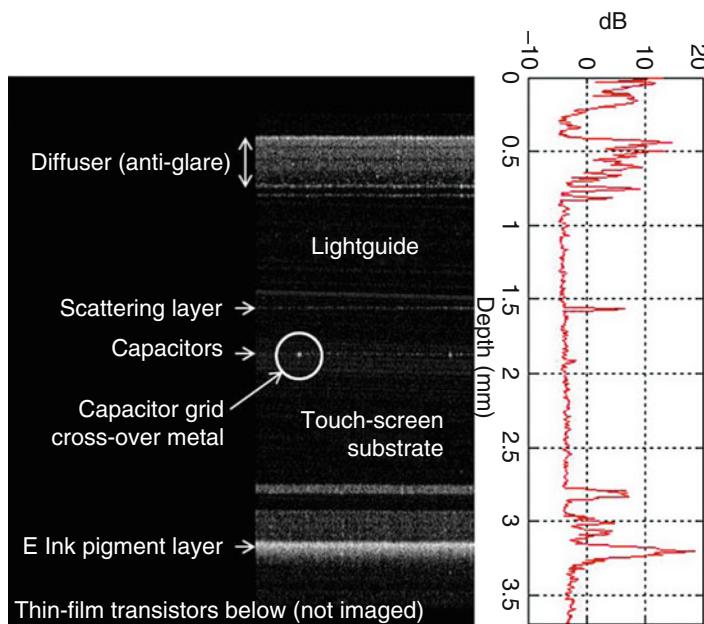
Besides the morphologic analysis of LEDs and solar cells, which allows the assessment of manufacturing defects in such layered devices, OCT can also obtain functional parameters under certain parameter-dependent conditions. In the case of high-power InGaN-based LEDs, it was demonstrated that OCT is suitable in determining the coefficients of thermal expansion depending on the injected currents as a

spatial distribution (Lee et al. 2017), which are pertinent parameters for the stability and reliability of LED devices. Therefore, cross-sectional images of the same sample were acquired under varying conditions and OCT offered the required time and spatial resolution to monitor the occurring expansion.

Comparable results, as for the application of OCT on solar cells and LEDs, have been presented in several publications on displays and will be mentioned by means of two exemplary studies. It was demonstrated that an OCT system in the 800 nm range is suitable for measurement of subsurface structures within the optical adhesives of a touch-screen panel (Cho et al. 2015), which allows the detection of both impurities in the adhesives and damages within individual layers and components. In a further approach, the reverse engineering, or so-called optical teardown, of a commercially available electrophoretic display, referred to as “e-paper,” was conducted by OCT with a swept source system in the 1060 nm range (Johnson et al. 2015). Besides the visualization of the incorporated layers for displaying and touch sensing that are presented in Fig. 12, phase-sensitive OCT was used to measure the particle motion of the display switching between black and white.

## Printed Electronics

Similar to adaptive manufacturing or 3D printing for workpieces, high-speed printing of electronic and photonic device has recently gained significant attention due to



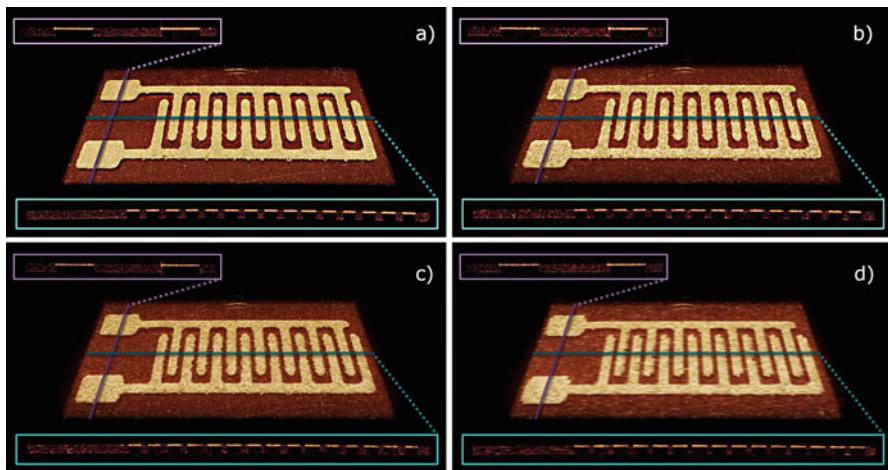
**Fig. 12** Cross-section (left) and single A-line (right) of a commercial e-paper display with corresponding components and layers by Swept Source OCT at 1060 nm. (Reprinted by permission from Johnson et al. 2015)

its potential to provide technically advanced fabricates at a low cost. While several printing techniques, including roll-to-roll processes (R2R), have been successfully demonstrated and implemented for the fabrication of flexible solar cells, LEDs, and other devices, there is a need for real-time and noninvasive inspection methods that provide structural properties as thickness, line width, homogeneity, and roughness in such processes. Czajkowski et al. were the first to apply an ultra-high-resolution TD-OCT system on RF antennas as an exemplary printed electronic device, whereby a supercontinuum generation process following a pump laser was utilized as a very broadband light source, offering a sub-micrometer axial resolution (Czajkowski et al. 2010). As single layers of such multilayered, multi-material products are usually on the order of a few up to tens of microns, achieving an axial resolution in the micron to submicron range is critical for the reliable evaluation of subsurface structures. In this case, information about the thickness of the encapsulation or surface protection layer and its uniformity might play a key role in a product's lifetime and allow for the correction and adjustment of the fabrication process to provide mechanical, moisture, and chemical protection. One drawback of the demonstrated TD-OCT system was the relatively low A-scan rate and therefore cross-sectional measurement speed, which was mainly limited by the axial scanning of the reference mirror. In a later study, the same light source was applied in a spectrometer-based FD-OCT setup (Czajkowski et al. 2012), enabling an A-scan rate of up to 140 kHz at a similarly high resolution, which is assumed to be sufficient for in-line monitoring in a roll-to-roll process. Feasibility was demonstrated by means of several samples including 3- $\mu\text{m}$ -thin layer of Parylene C, an organic solar cell, and a printed battery electrode.

Based on previous results, the capability of OCT for the integration in a real manufacturing procedure was subsequently shown in a simulated roll-to-roll process at an A-scan rate of 127 kHz (Alarousu et al. 2013). Internal structures at a micron level of a moving printed electrode could be resolved at speeds of up to 1 m/min and with minor limitations up to 1.5 m/min, as it is shown in Fig. 13. Remaining limitations of the applied system were therefore mentioned to be the camera acquisition rate, beam deflection mechanics, and the handling and processing of the large amount of acquired data. As these issues have already been solved elsewhere, OCT is an applicable technique for nondestructive testing in printing processes for electronic devices.

## Optical Surfaces and Samples

The manufacturing advancements of optical systems are a decisive drive factor in the recent development of OCT systems, as it is, for instance, apparent in the demonstration of novel endoscopic optics that are based on gradient-index (GRIN) lenses. But also the emerging developments of new types of optics themselves, including freeform surfaces that are characterized by their nonrotational symmetry, have driven the need of enhanced metrology systems that provide spatially resolved information about surface and subsurface structures as well as refractive indices and material



**Fig. 13** 3D reconstruction of the printed interdigitated electrodes moving with varying speeds. (a) Stationary sample. (b) Sample moving at a speed of 0.45 m/min, (c) at 0.75 m/min, and (d) at 1.50 m/min. Reconstructions differ in sampling resolution due to a constant scanning speed. Violet and turquoise boxes represent slice images in cross direction and machine direction, respectively. The slice images were taken from the actual 3D data and not measured separately. (Reprinted by permission from Springer Customer Service Centre GmbH: Springer Nature, Alarousu et al. 2013)

parameters. As OCT has been demonstrated as a valuable tool for the assessment of such parameters, an overview of recent applications is given here.

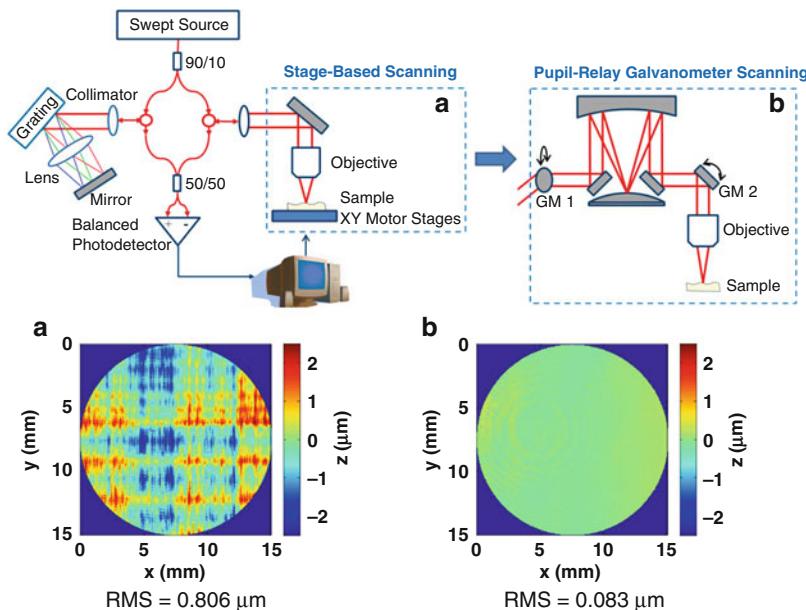
## Surfaces

The evaluation of surface and subsurface structures is the main application of OCT in the field of nondestructive testing. In the case of optical surfaces, conventional interferometric techniques, based on relative distance measurements, can be easily applied for the measurement of polished optical surfaces, but grinded surfaces, which typically occur as an intermediate step in the fabrication of aspherical and free-from glass surfaces, exhibit the limitations of those approaches as the surface roughness commonly exceeds the systems' working parameter. Alternatively, tactile measurements can be applied at the residual risk of altering the sample surface by the measurement itself. OCT as a noncontact method, therefore, has been demonstrated for the measurement of aspheric grinded surfaces (Hellmuth et al. 2007) and provides several advantages compared to conventional methods, in particular, by the detection of subsurface damages in addition to surface roughness parameters. Comparable results were obtained in a study that utilized OCT for in situ monitoring of surfaces in large-aperture fused silica optics at a long-working-distance geometry, allowing the OCT imaging through the opposite surface of the 5-cm-thick sample (Guss et al. 2008).

For automated surface inspections in an uncontrolled environment, fake defects such as dust or particles challenge the discrimination of real surface defects on the optical sample, even if three-dimensional structural data are provided. A recent study

has addressed this issue by analyzing defect peripheries with a parallel spectral-domain OCT system in the 800 nm wavelength range (Chen et al. 2015) that utilized surface deformations in the demarcation area of the defect, which were mapped by a phase-sensitive evaluation of the complex OCT signal. The study therefore demonstrated the discrimination of embedded defects and dust particles with sizes down to 100  $\mu\text{m}$  on a prepared glass sample with remaining potential for improvements, in particular of the lateral resolution (15.6  $\mu\text{m}$ ) in the applied system as it was mentioned by the authors.

More complex and nonsymmetric surfaces, which appear in the manufacturing of optical free-form components, have recently driven the demand for high-performance optical metrology tools that are capable of significant slopes in the examined parts due to improved sensitivity. To achieve reliable results with a scanning approach, a custom pupil-relay scanning configuration, as displayed in Fig. 14, was presented replacing the translational stage-based scanning, which reduced the RMS discrepancy of the applied swept source system from 0.8  $\mu\text{m}$  to 0.08  $\mu\text{m}$  for an optical flat standard (Yao et al. 2015b). The further development of this approach led recently to the detailed description of a point-cloud noncontact metrology system (Yao et al. 2018), which included a comprehensive set of reconstructions, rendering and fitting algorithms to evaluate the metrology results and investigate the uncertainty in the measurements.



**Fig. 14** Upper part: Swept source OCT freeform metrology system layout with Galvanometer mirrors (GM) and different scanning approaches. Lower part: Measured Surface Profile of an optical flat standard using (a) the stage-based scanning and (b) the pupil-relay galvanometer scanning system (Reprinted by permission from SPIE Publications, Yao et al. 2015b)

## Gradient-Index Optics

Offering a gradual variation of the refractive index within the optic component, gradient-index (GRIN) optics enable the production of lenses with flat surfaces and reduced aberrations that can be mounted in a miniaturized lens system. Present manufacturing of GRIN optics is mainly limited by the maximum gradient that can be realized in the refractive index profile, by the component volume, as only diameters in the mm range are fabricated cost-effectively, and in the inability to combine freeform surfaces with arbitrary gradients.

Recently, the manufacturing has advanced significantly by compositing polymer nanolayers with individual refractive indices into a consolidated GRIN sheet, which eventually may enable components of arbitrary shape and gradient profile. As the assessment of the refractive index profile is not only important for the performance evaluation in optical devices but in particular of interest for the quality control of such products, several techniques have been proposed targeting those parameters based on ray tracing, for instance. In the case of a GRIN rod lens, it was demonstrated that OCT enables the nondestructive measurement of the refractive index profile (Lv et al. 2010), whereby the optical paths of the light traveling through the rod lens in radial direction were evaluated.

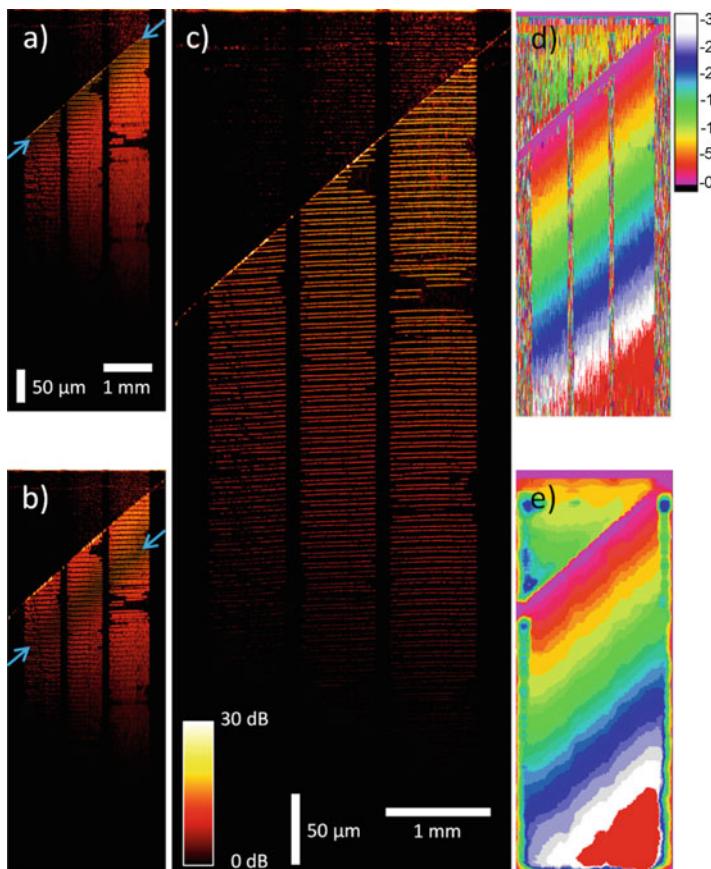
A more in-depth examination of the metrology of GRIN optics was presented in several publications by a group from Rochester (NY) that utilized OCT to quantify characteristics of layered polymeric GRIN material, so-called GRIN sheets, aiming to provide a direct feedback into the fabrication process (Meemon et al. 2013). As the proposed technique was mainly applied to flat surfaces, the translation toward curved surfaces required a modification of the scanning scheme to overcome the penetration loss for incident angles far apart from the normal incident angles. This was addressed by an angular scanning scheme (Yao et al. 2015a) for the 3D metrology of spherical GRIN preforms, which allows mapping of the group refractive index, physical thickness, and the transmitted wave front. Later, those results were utilized to assess internal defects and misalignments of the index distribution of a layered spherical GRIN lens, which were successfully exploited to achieve significant spherical aberration reduction of such lenses and increased effective aperture (Yao et al. 2016). Additionally, it was pointed out by the authors that the developed system and concept is also applicable to render general freeform samples with potentially arbitrary shapes.

## Ferroelectric Media

Ferroelectric crystals with a periodic domain reversal, such as lithium niobate ( $\text{LiNbO}_3$ ) and lithium tantalate ( $\text{LiTaO}_3$ ), are of great interest due to their widespread applications in wavelength conversion, for instance, frequency doubling and electro-optics. To fully characterize the domain structure of periodically poled lithium niobate (PPLN) and the related domain walls noninvasively, OCT was first utilized by Pei et al. in a study that visualized domain walls in PPLN in the extraordinary polarization (Pei et al. 2011). As the axial resolving power of an OCT system increases by the factor of the sample's refractive index, a value of about 2.2 for the extraordinary

polarization facilitates in this case submicron resolution of structures within the sample. In addition, the samples' high dispersion in the spectral range around 700 nm and, thus apparently, dispersion broadening of single reflexes in the measurement only allowed for a visualization of the first domain walls in depth, while the characterization of the observed dispersion was in good agreement with theory.

To overcome this limitation, a dispersion compensation concept based on Fourier Domain OCT was introduced by Kirsten et al., which significantly improved the ability of OCT to assess the domain structure over a depth range of several hundred  $\mu\text{m}$  (Kirsten et al. 2017), as is shown in Fig. 15. Thus, the approach yielded an accuracy of better than 10 nm to measure the average period of the domains. Furthermore, the ratio



**Fig. 15** Image of the PPLN crystal with (a) optimal dispersion compensation for the surface of the crystal, (b) additional dispersion compensation of  $5\pi$ , in both images the stripe with optimal dispersion is indicated by blue arrows. (c) Overlay of parts with optimal dispersion compensation. (d) Color map of optimal dispersion correction. The additional phase correction has an amplitude corresponding to the color scale in units of  $\pi$ . (e) The dispersion correction is smoothed horizontally to have reasonable values in areas without signal. (Scale bar calculated for PPLN area.). All images show the same area. (Reprinted with permission from Kirsten et al. 2017, OSA Publishing)

between domain thickness and period could be evaluated and the fan-shaped structure of a variable frequency doubler was visualized. Besides the static mapping of ferroelectric domain walls, the same system was employed to dynamically measure the domain structure under external fields (Haußmann et al. 2017), where the electro-optic refractive index dependency was assessed. There, a stable OCT signal of the domain walls up to applied fields at the switching threshold was observed, whereas domain wall movements for higher fields could be demonstrated, leaving the potential for a deeper understanding of such dynamical processes by OCT.

All three studies indicate that measurements of the domain structure of PPLN crystals by ultrahigh-resolution OCT can potentially be used to optimize the fabrication and nondestructive assessment of quality and imperfections of a certain crystal.

---

## Summary

In this chapter, the technique of OCT has been described first. System parameters, limits, and extensions of the method have been briefly illustrated with references for the more interested reader. Although the technology of OCT has advanced tremendously in many ways during the last 25 years, this technique is still rarely used in NDE/NDT. The examples provided in this chapter show the potential of this method in a wide range of applications outside the biomedical field, where it has become a clinical standard in ophthalmology and intravascular imaging and is emerging into other fields. One reason for the rare use of OCT in NDE might be that there are no commercial turnkey instruments for typical applications on the market yet, so that specialized personnel is needed to set up an instrument for a specific application. But universal instruments, which can be tailored to many applications, are on the market, making the application much more easy than starting from scratch. Therefore, the use of OCT in NDE will probably grow rapidly during the next years.

---

## References

- Adler DC, Stenger J, Gorczynska I, Lie H, Hensick T, Spronk R, Wolohojian S, Khandekar N, Jiang JY, Barry S (2007) Comparison of three-dimensional optical coherence tomography and high resolution photography for art conservation studies. *Opt Express* 15(24):15972–15986
- Akcay AC, Rolland JP, Eichenholz JM (2003) Spectral shaping to improve the point spread function in optical coherence tomography. *Opt Lett* 28(20):1921–1923
- Akhondi E, Wu B, Sun S, Marxer B, Lim W, Gu J, Liu L, Burkhardt M, McDougald D, Pronk W, Fane AG (2015) Gravity-driven membrane filtration as pretreatment for seawater reverse osmosis: linking biofouling layer morphology with flux stabilization. *Water Res* 70:158–173. <https://doi.org/10.1016/j.watres.2014.12.001>
- Alarousu E, AlSaggaf A, Jabbour GE (2013) Online monitoring of printed electronics by spectral-domain optical coherence tomography. *Sci Rep* 3:1562. <https://doi.org/10.1038/srep01562>
- Antoniuk P, Strąkowski M, Pluciński J, Kosmowski B (2012) Non-destructive inspection of anti-corrosion protective coatings using optical coherent tomography. *Metrology Measur Syst* 19(2):365–372

- Bail MA, Haeusler G, Herrmann JM, Lindner MW, Ringler R (1996) Optical coherence tomography with the “spectral radar”: fast optical analysis in volume scatterers by short-coherence interferometry. In: Benaron DA, Chance B, Mueller GJ (eds). In Photon Propagation in Tissues II (Vol. 2925, pp. 298–303)
- Bemand E, Liang H (2013) Optical coherence tomography for vulnerability assessment of sandstone. *Appl Opt* 52(14):3387–3393. <https://doi.org/10.1364/AO.52.003387>
- Buchsbaum A, Egger M, Burzic I, Koepllmayr T, Aigner M, Miethlinger J, Leitner M (2015) Optical coherence tomography based particle image velocimetry (OCT-PIV) of polymer flows. *Opt Lasers Eng* 69:40–48
- Burkhardt A, Geissler S, Koch E (2010) Optical coherence tomography as approach for the minimal invasive localization of the germinal disc in ovo before chicken sexing. In Biophotonics: Photonic Solutions for Better Health Care II (Vol. 7715)
- Burkhardt A, Kirsten L, Bornitz M, Zahnert T, Koch E (2014) Investigation of the human tympanic membrane oscillation ex vivo by Doppler optical coherence tomography. *J Biophotonics* 7(6):434–441
- Campello SL, dos Santos WP, Machado VF, Mota CCBO, Gomes ASL, de Souza RE (2014) Microstructural information of porous materials by optical coherence tomography. *Microporous Mesoporous Mater* 198:50–54. <https://doi.org/10.1016/j.micromeso.2014.07.009>
- Chang S, Mao Y, Chang G, Flueraru C (2010) Jade detection and analysis based on optical coherence tomography images. *Optical Engineering*, 49(6)
- Chen Z, Shen Y, Bao W, Li P, Wang X, Ding Z (2015) Identification of surface defects on glass by parallel spectral domain optical coherence tomography. *Opt Express* 23(18):23634–23646. <https://doi.org/10.1364/OE.23.023634>
- Cheung CS, Tokurakawa M, Daniel JMO, Clarkson WA, Liang H (2013) Long wavelength optical coherence tomography for painted objects. In Optics for Arts, Architecture, and Archaeology IV (Vol. 8790)
- Cho NH, Jung U, Kim S, Kim J (2012) Non-destructive inspection methods for LEDs using real-time displaying optical coherence tomography. *Sensors* 12(8):10395
- Cho NH, Park K, Kim J-Y, Jung Y, Kim J (2015) Quantitative assessment of touch-screen panel by nondestructive inspection with three-dimensional real-time display optical coherence tomography. *Opt Lasers Eng* 68:50–57. <https://doi.org/10.1016/j.optlaseng.2014.12.013>
- Choi W-J, Min G-H, Lee B-H, Eom J-H, Kim J-W (2010) Counterfeit detection using characterization of safety feature on banknote with full-field optical coherence tomography. *J Opt Soc Korea* 14(4):316–320
- Choi WJ, Jung SP, Shin JG, Yang D, Lee BH (2011) Characterization of wet pad surface in chemical mechanical polishing (CMP) process with full-field optical coherence tomography (FF-OCT). *Opt Express* 19(14):13343–13350. <https://doi.org/10.1364/OE.19.013343>
- Choma M, Sarunic M, Yang C, Izatt J (2003) Sensitivity advantage of swept source and Fourier domain optical coherence tomography. *Opt Express* 11(18):2183–2189
- Choma MA, Hsu K, Izatt JA (2005) Swept source optical coherence tomography using an all-fiber 1300-nm ring laser source. *J Biomed Opt* 10(4):044009
- Cimalla P, Walther J, Mehner M, Cuevas M, Koch E (2009) Simultaneous dual-band optical coherence tomography in the spectral domain for high resolution in vivo imaging. *Opt Express* 17(22):19486–19500
- Czajkowski J, Prykäri T, Alarousu E, Palosaari J, Myllylä R (2010) Optical coherence tomography as a method of quality inspection for printed electronics products. *Opt Rev* 17(3):257–262. <https://doi.org/10.1007/s10043-010-0045-0>
- Czajkowski J, Lauri J, Sliz R, Fält P, Fabritius T, Myllylä R, Cense B (2012) Sub-micron resolution high-speed spectral domain optical coherence tomography in quality inspection for printed electronics. In Optical Micro-and Nanometrology IV (Vol. 8430)
- de Boer JF, Cense B, Park BH, Pierce MC, Tearney GJ, Bouma BE (2003) Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography. *Opt Lett* 28(21):2067–2069

- de Boer JF, Hitzenberger CK, Yasuno Y (2017) Polarization sensitive optical coherence tomography – a review [invited]. *Biomed Opt Express* 8(3):1838–1873. <https://doi.org/10.1364/Boe.8.001838>
- Dingemans LM, Papadakis VM, Liu P, Adam AJL, Groves RM (2015) Optical coherence tomography complemented by hyperspectral imaging for the study of protective wood coatings. In: *Optics for Arts, Architecture, and Archaeology V* (Vol. 9527)
- Dong B, Chen S, Zhou F, Chan CHY, Yi J, Zhang HF, Sun C (2016) Real-time functional analysis of inertial microfluidic devices via spectral domain optical coherence tomography. *Sci Rep* 6:33250. <https://doi.org/10.1038/srep33250>
- Dong Y, Lin H, Abolghasemi V, Gan L, Zeitler JA, Shen Y-C (2017) Investigating intra-tablet coating uniformity with spectral-domain optical coherence tomography. *J Pharm Sci* 106(2):546–553. <https://doi.org/10.1016/j.xphs.2016.09.021>
- Dreszer C, Wexler AD, Drusová S, Overdijk T, Zwijnenburg A, Flemming HC, Kruithof JC, Vrouwenvelder JS (2014) In-situ biofilm characterization in membrane systems using optical coherence tomography: formation, structure, detachment and impact of flux change. *Water Res* 67:243–254. <https://doi.org/10.1016/j.watres.2014.09.006>
- Drexler W, Morgner U, Kartner FX, Pitris C, Boppart SA, Li XD, Ippen EP, Fujimoto JG (1999) In vivo ultrahigh-resolution optical coherence tomography. *Opt Lett* 24(17):1221–1223
- Dunkers JP, Parnas RS, Zimba CG, Peterson RC, Flynn KM, Fujimoto JG, Bouma BE (1999) Optical coherence tomography of glass reinforced polymer composites. *Compos A: Appl Sci Manuf* 30(2):139–145. [https://doi.org/10.1016/S1359-835X\(98\)00084-0](https://doi.org/10.1016/S1359-835X(98)00084-0)
- Fan C, Yao G (2012) 3D imaging of tomato seeds using frequency domain optical coherence tomography. In: *Sensing for Agriculture and Food Quality and Safety IV* (Vol. 8369)
- Felton LA (2016) Aqueous polymeric coatings for pharmaceutical dosage forms. CRC Press, Boca Raton
- Fercher AF, Mengedoht K, Werner W (1988) Eye-length measurement by interferometry with partially coherent light. *Opt Lett* 13(3):186–188
- Fercher AF, Hitzenberger CK, Kamp G, El-Zaiat SY (1995) Measurement of intraocular distances by backscattering spectral interferometry. *Opt Commun* 117(1–2):43–48. [https://doi.org/10.1016/0030-4018\(95\)00119-S](https://doi.org/10.1016/0030-4018(95)00119-S)
- Fercher AF, Hitzenberger CK, Sticker M, Zawadzki R, Karamata B, Lasser T (2002) Dispersion compensation for optical coherence tomography depth-scan signals by a numerical technique. *Opt Commun* 204(1–6):67–74
- Ford HD, Tatam RP (2013) Spatially-resolved volume monitoring of adhesive cure using correlated-image optical coherence tomography. *Int J Adhes Adhes* 42:21–29. <https://doi.org/10.1016/j.ijadhadh.2012.11.012>
- Fortunato L, Leiknes T (2017) In-situ biofouling assessment in spacer filled channels using optical coherence tomography (OCT): 3D biofilm thickness mapping. *Bioresour Technol* 229:231–235. <https://doi.org/10.1016/j.biortech.2017.01.021>
- Fuchs S, Wusnsche M, Nathanael J, Abel JJ, Rodel C, Biedermann J, Reinhard J, Hubner U, Paulus GG (2017) Optical coherence tomography with nanoscale axial resolution using a laser-driven high-harmonic source. *Optica* 4(8):903–906. <https://doi.org/10.1364/Optica.4.000903>
- Fujiwara K, Matoba O (2011a) Detection and evaluation of security features embedded in paper using spectral-domain optical coherence tomography. *Opt Rev* 18(1):171–175. <https://doi.org/10.1007/s10043-011-0018-y>
- Fujiwara K, Matoba O (2011b) High-speed cross-sectional imaging of valuable documents using common-path swept-source optical coherence tomography. *Appl Opt* 50(34):H165–H170. <https://doi.org/10.1364/AO.50.00H165>
- Gandhi V, Semenov D, Honkanen S, Hauta-Kasari M (2015) Optical identification based on time domain optical coherence tomography. *Appl Opt* 54(25):7514–7519. <https://doi.org/10.1364/AO.54.007514>
- Gardner MR, Lewis A, Park J, McElroy AB, Estrada AD, Fish S, Beaman JJ, Milner TE (2018) In situ process monitoring in selective laser sintering using optical coherence tomography. *Optical Engineering*, 57(4)

- Gastinger K, Johnsen L, Simonsen O, Aksnes A (2011) Inspection of processes during silicon wafer sawing using low coherence interferometry in the near infrared wavelength region. In Optical Measurement Systems for Industrial Inspection VII (Vol. 8082)
- Goldberg GR, Boldin A, Andersson SML, Ivanov P, Ozaki N, Taylor RJE, Childs DTD, Groom KM, Kennedy KL, Hogg RA (2017) Gallium nitride Superluminescent light emitting diodes for optical coherence tomography applications. *IEEE J Sel Top Quantum Electron* 23(6):1–11. <https://doi.org/10.1109/JSTQE.2017.2732941>
- Gora M, Targowski P, Rycyk A, Marczak J (2006) Varnish ablation control by optical coherence tomography. *Laser Chem* 2006:10647
- Götzinger E, Pircher M, Geitzenauer W, Ahlers C, Baumann B, Michels S, Schmidt-Erfurth U, Hitzenberger CK (2008) Retinal pigment epithelium segmentation by polarization sensitive optical coherence tomography. *Opt Express* 16(21):16410–16422. <https://doi.org/10.1364/OE.16.016410>
- Grombe R, Kirsten L, Mehner M, Linsinger TPJ, Emons H, Koch E (2014) Feasibility of non-invasive detection of engineered nanoparticles in food mimicking matrices by optical coherence tomography. *Food Chem* 153:444–449. <https://doi.org/10.1016/j.foodchem.2013.12.089>
- Grombe R, Kirsten L, Mehner M, Linsinger TPJ, Koch E (2016) Improved non-invasive optical coherence tomography detection of different engineered nanoparticles in food-mimicking matrices. *Food Chem* 212:571–575. <https://doi.org/10.1016/j.foodchem.2016.06.008>
- Guan G, Hirsch M, Syam WP, Leach RK, Huang Z, Clare AT (2016) Loose powder detection and surface characterization in selective laser sintering via optical coherence tomography. *Proc R Soc A Math Phys Eng Sci* 472(2191):201607. <https://doi.org/10.1098/rspa.2016.0201>
- Guss GM, Bass IL, Hackel RP, Mailhot C, Demos SG (2008) In situ monitoring of surface postprocessing in large-aperture fused silica optics with optical coherence tomography. *Appl Opt* 47(25):4569–4573. <https://doi.org/10.1364/AO.47.004569>
- Haindl R, Träischker W, Wartak A, Baumann B, Pircher M, Hitzenberger CK (2016) Total retinal blood flow measurement by three beam Doppler optical coherence tomography. *Biomed Opt Express* 7(2):287–301
- Haußmann A, Kirsten L, Schmidt S, Cimalla P, Wehmeier L, Koch E, Eng LM (2017) Three-dimensional, time-resolved profiling of ferroelectric domain wall dynamics by spectral-domain optical coherence tomography. *Ann Phys* 529(8):1700139. <https://doi.org/10.1002/andp.201700139>
- Heise B, Wiesauer K, Götzinger E, Pircher M, Hitzenberger CK, Engelke R, Ahrens G, Grützner G, Stifter D (2010) Spatially resolved stress measurements in materials with polarisation-sensitive optical coherence tomography: image acquisition and processing aspects. *Strain* 46(1):61–68. <https://doi.org/10.1111/j.1475-1305.2008.00589.x>
- Hellmuth T, Börret R, Khrennikov K (2007) 3-dimensional scanning of grinded optical surfaces based on optical coherence tomography. In: Optical engineering + applications. In Optical Manufacturing and Testing VII (Vol. 6671, p. 66710X). International Society for Optics and Photonics
- Hierzenberger P, Leiss-Holzinger E, Heise B, Stifter D, Eder G (2014) In-situ optical coherence tomography (OCT) for the time-resolved investigation of crystallization processes in polymers. *Macromolecules* 47(6):2072–2079. <https://doi.org/10.1021/ma4023839>
- Hitzenberger CK (1991) Optical measurement of the axial eye length by laser Doppler interferometry. *Invest Ophthalmol Vis Sci* 32(3):616–624
- Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, Hee MR, Flotte T, Gregory K, Puliafito CA, Fujimoto JG (1991) Optical coherence tomography. *Science* 254 (5035):1178–1181
- Huber R, Adler DC, Fujimoto JG (2006) Buffered Fourier domain mode locking: unidirectional swept laser sources for optical coherence tomography imaging at 370,000 lines/s. *Opt Lett* 31(20):2975–2977
- Hughes M, Spring M, Podoleanu A (2010) Speckle noise reduction in optical coherence tomography of paint layers. *Appl Opt* 49(1):99–107

- Johnson B, Atia W, Kuznetsov M, Larson N, McKenzie E, Mathur V, Goldberg B, Whitney P (2015) Optical teardown of a Kindle Paperwhite display by OCT. arXiv preprint arXiv:160505174
- Ju MJ, Lee SJ, Min EJ, Kim Y, Kim HY, Lee BH (2010) Evaluating and identifying pearls and their nuclei by using optical coherence tomography. *Opt Express* 18(13):13468–13477
- Kang J, Feng P, Wei X, Lam EY, Tsia KK, Wong KKY (2018) 102-nm, 44.5-MHz inertial-free swept source by mode-locked fiber laser and time stretch technique for optical coherence tomography. *Opt Express* 26(4):4370–4381. <https://doi.org/10.1364/OE.26.004370>
- Kirillin MY, Myllylä R, Priezzhev AV (2007) Optical coherence tomography of paper: Monte Carlo simulation for multilayer model. In Saratov Fall Meeting 2006: Coherent Optics of Ordered and Random Media VII (Vol. 6536)
- Kirillin MY, Priezzhev AV, Myllylä R (2008) Effect of paper porosity on OCT images: Monte Carlo study. In Advanced Laser Technologies 2007 (Vol. 7022)
- Kirsten L, Domaschke T, Schneider C, Walther J, Meissner S, Hampel R, Koch E (2015) Visualization of dynamic boiling processes using high-speed optical coherence tomography. *Exp Fluids* 56(3):52. <https://doi.org/10.1007/s00348-015-1921-y>
- Kirsten L, Haußmann A, Schnabel C, Schmidt S, Cimalla P, Eng LM, Koch E (2017) Advanced analysis of domain walls in Mg doped LiNbO<sub>3</sub> crystals with high resolution OCT. *Opt Express* 25(13):14871–14882. <https://doi.org/10.1364/OE.25.014871>
- Koller DM, Hanneschläger G, Leitner M, Khinast JG (2011) Non-destructive analysis of tablet coatings with optical coherence tomography. *Eur J Pharm Sci* 44(1):142–148. <https://doi.org/10.1016/j.ejps.2011.06.017>
- Kottig F, Cimalla P, Gartner M, Koch E (2012) An advanced algorithm for dispersion encoded full range frequency domain optical coherence tomography. *Opt Express* 20(22):24925–24948
- Latour G, Echar J-P, Soulier B, Emond I, Vaiedelich S, Elias M (2009) Structural and optical properties of wood and wood finishes studied using optical coherence tomography: application to an 18th century Italian violin. *Appl Opt* 48(33):6485–6491. <https://doi.org/10.1364/AO.48.006485>
- Lee BS, Strand TC (1990) Profilometry with a coherence scanning microscope. *Appl Opt* 29(26):3784–3788. <https://doi.org/10.1364/AO.29.003784>
- Lee C, Lee S-Y, Kim J-Y, Jung H-Y, Kim J (2011) Optical sensing method for screening disease in melon seeds by using optical coherence tomography. *Sensors* 11(10):9467
- Lee S-Y, Lee C, Kim J, Jung H-Y (2012) Application of optical coherence tomography to detect cucumber green mottle mosaic virus (CGMMV) infected cucumber seed. *Hortic Environ Biotechnol* 53(5):428–433. <https://doi.org/10.1007/s13580-012-0071-x>
- Lee Y-J, Chou C-Y, Huang C-Y, Yao Y-C, Haung Y-K, Tsai M-T (2017) Determination on the coefficient of thermal expansion in high-power InGaN-based light-emitting diodes by optical coherence tomography. *Sci Rep* 7(1):14390. <https://doi.org/10.1038/s41598-017-14689-y>
- Lei M, Sun Y, Wang D, Li P (2009) Automated thickness measurements of pearl from optical coherence tomography images. In: IEEE, pp 247–251
- Leitgeb R, Wojtkowski M, Kowalczyk A, Hitzenberger CK, Sticker M, Fercher AF (2000) Spectral measurement of absorption by spectroscopic frequency-domain optical coherence tomography. *Opt Lett* 25(11):820–822
- Leitgeb R, Hitzenberger C, Fercher A (2003a) Performance of fourier domain vs. time domain optical coherence tomography. *Opt Express* 11(8):889–894
- Leitgeb RA, Hitzenberger CK, Fercher AF, Bajraszewski T (2003b) Phase-shifting algorithm to achieve high-speed long-depth-range probing by frequency-domain optical coherence tomography. *Opt Lett* 28(22):2201–2203. <https://doi.org/10.1364/OL.28.002201>
- Lenz M, Mazzon C, Dillmann C, Gerhardt NC, Welp H, Prange M, Hofmann MR (2017) Spectral domain optical coherence tomography for non-destructive testing of protection coatings on metal substrates. *Appl Sci* 7(4):364
- Lewis AD, Katta N, McElroy AB, Milner TE, Fish S, Beaman JJ (2018) Understanding and improving optical coherence tomography imaging depth in selective laser sintering nylon 12 parts and powder. *Optical Engineering*, 57(4), 041414

- Li M, Verboven P, Buchsbaum A, Cantre D, Nicolaï B, Heyes J, Mowat A, East A (2015) Characterising kiwifruit (*Actinidia* sp.) near skin cellular structures using optical coherence tomography. Postharvest Biol Technol 110:247–256. <https://doi.org/10.1016/j.postharvbio.2015.08.021>
- Li C, Felz S, Wagner M, Lackner S, Horn H (2016) Investigating biofilm structure developing on carriers from lab-scale moving bed biofilm reactors based on light microscopy and optical coherence tomography. Bioresour Technol 200:128–136. <https://doi.org/10.1016/j.biortech.2015.10.013>
- Liang H, Lange R, Peric B, Spring M (2013) Optimum spectral window for imaging of art with optical coherence tomography. Appl Phys B 111(4):589–602
- Liang H, Cheung CS, Daniel JMO, Tokurakawa M, Clarkson WA (2015) Spring M High resolution Fourier domain Optical Coherence Tomography at 2 microns for painted objects. In: International society for optics and photonics, p 952705
- Lin H, Dong Y, Shen Y, Axel Zeitler J (2015) Quantifying pharmaceutical film coating with optical coherence tomography and terahertz pulsed imaging: an evaluation. J Pharm Sci 104(10): 3377–3385. <https://doi.org/10.1002/jps.24535>
- Lin H, Dong Y, Markl D, Williams BM, Zheng Y, Shen Y, Zeitler JA (2017a) Measurement of the intertablet coating uniformity of a pharmaceutical pan coating process with combined terahertz and optical coherence tomography in-line sensing. J Pharm Sci 106(4):1075–1084. <https://doi.org/10.1016/j.xphs.2016.12.012>
- Lin H, Dong Y, Markl D, Zhang Z, Shen Y, Zeitler JA (2017b) Pharmaceutical film coating catalog for spectral domain optical coherence tomography. J Pharm Sci 106(10):3171–3176. <https://doi.org/10.1016/j.xphs.2017.05.032>
- Lippok N, Coen S, Nielsen P, Vanholsbeeck F (2012) Dispersion compensation in Fourier domain optical coherence tomography using the fractional Fourier transform. Opt Express 20(21): 23398–23413
- Liu X, Kang JU (2010) Compressive SD-OCT: the application of compressed sensing in spectral domain optical coherence tomography. Opt Express 18(21):22010–22019. <https://doi.org/10.1364/OE.18.022010>
- Liu C, Wong A, Bizheva K, Fieguth P, Bie H (2012) Homotopic, non-local sparse reconstruction of optical coherence tomography imagery. Opt Express 20(9):10200–10211. <https://doi.org/10.1364/OE.20.010200>
- Liu J, Tian XL, Sun YK (2013) Pearl thickness measurements from optical coherence tomography images. In: Trans Tech Publ, pp 415–420
- Liu P, Groves RM, Benedictus R (2014a) 3D monitoring of delamination growth in a wind turbine blade composite using optical coherence tomography. NDT & E International 64:52–58. <https://doi.org/10.1016/j.ndteint.2014.03.003>
- Liu P, Groves RM, Benedictus R (2014b) Non-destructive evaluation of delamination growth in glass fiber composites using optical coherence tomography. In: Nondestructive Characterization for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security 2014 (Vol. 9063, p. 90631M). International Society for Optics and Photonics
- Liu P, Groves RM, Benedictus R (2014c) Optical coherence elastography for measuring the deformation within glass fiber composite. Appl Opt 53(22):5070–5077. <https://doi.org/10.1364/AO.53.005070>
- Liu P, Groves RM, Benedictus R (2014d) Optical coherence tomography for the study of polymer and polymer matrix composites. Strain 50(5):436–443. <https://doi.org/10.1111/str.12095>
- Liu X, Li W, Chong TH, Fane AG (2017a) Effects of spacer orientations on the cake formation during membrane fouling: quantitative analysis based on 3D OCT imaging. Water Res 110:1–14. <https://doi.org/10.1016/j.watres.2016.12.002>
- Liu X, Zaki F, Wang Y, Huang Q, Mei X, Wang J (2017b) Secure fingerprint identification based on structural and microangiographic optical coherence tomography. Appl Opt 56(8):2255–2259. <https://doi.org/10.1364/AO.56.002255>
- Lorenser D, Singe CC, Curatolo A, Sampson DD (2014) Energy-efficient low-Fresnel-number Bessel beams and their application in optical coherence tomography. Opt Lett 39(3):548–551. <https://doi.org/10.1364/OL.39.000548>

- Lv H, Liu A, Tong J, Yi X, Li Q, Wang X, Ding Y (2010) Nondestructive measurement of refractive index profile of gradient refractive index rod lens. *Rev Sci Instrum* 81(10):103104. <https://doi.org/10.1063/1.3492154>
- Magwaza LS, Ford HD, Cronje PJR, Opara UL, Landahl S, Tatam RP, Terry LA (2013) Application of optical coherence tomography to non-destructively characterise rind breakdown disorder of 'Nules Clementine' mandarins. *Postharvest Biol Technol* 84:16–21. <https://doi.org/10.1016/j.postharvbio.2013.03.019>
- Manukyan S, Sauer HM, Roisman IV, Baldwin KA, Fairhurst DJ, Liang H, Venzmer J, Tropea C (2013) Imaging internal flows in a drying sessile polymer dispersion drop using spectral radar optical coherence tomography (SR-OCT). *J Colloid Interface Sci* 395:287–293. <https://doi.org/10.1016/j.jcis.2012.11.037>
- Markl D, Hanneschläger G, Sacher S, Leitner M, Khinast JG (2014) Optical coherence tomography as a novel tool for in-line monitoring of a pharmaceutical film-coating process. *Eur J Pharm Sci* 55:58–67. <https://doi.org/10.1016/j.ejps.2014.01.011>
- Markl D, Hanneschläger G, Sacher S, Leitner M, Khinast JG, Buchsbaum A (2015a) Automated pharmaceutical tablet coating layer evaluation of optical coherence tomography images. *Meas Sci Technol* 26(3):035701
- Markl D, Zettl M, Hanneschläger G, Sacher S, Leitner M, Buchsbaum A, Khinast JG (2015b) Calibration-free in-line monitoring of pellet coating processes via optical coherence tomography. *Chem Eng Sci* 125:200–208. <https://doi.org/10.1016/j.ces.2014.05.049>
- Meemon P, Yao J, Lee K-S, Thompson KP, Ponting M, Baer E, Rolland JP (2013) Optical coherence tomography enabling non destructive metrology of layered polymeric GRIN material. *Sci Rep* 3:1709. <https://doi.org/10.1038/srep01709>
- Meissner S, Breithaupt R, Koch E (2013) Defense of fake fingerprint attacks using a swept source laser optical coherence tomography setup. In: *Frontiers in Ultrafast Optics: Biomedical, Scientific, and Industrial Applications XIII* (Vol. 8611, p. 86110L). International Society for Optics and Photonics
- Pan L, Wang X, Li Z, Zhang X, Bu Y, Nan N, Chen Y, Wang X (2017) Depth-dependent dispersion compensation for full-depth OCT image. *Opt Express* 25(9):10345–10354
- Park S, You J, Ahn Y, Jung W, Kim J, Lee S, Park J, Cho KH (2018) Evaluating the effects of organic matter bioavailability on nanofiltration membrane using real-time monitoring. *J Membr Sci* 548:519–525. <https://doi.org/10.1016/j.memsci.2017.11.053>
- Pei S-C, Ho T-S, Tsai C-C, Chen T-H, Ho Y, Huang P-L, Kung AH, Huang S-L (2011) Non-invasive characterization of the domain boundary and structure properties of periodically poled ferroelectrics. *Opt Express* 19(8):7153–7160. <https://doi.org/10.1364/OE.19.007153>
- Pfeiffer T, Draxinger W, Grill C, Huber R (2017) Long-range live 3D-OCT at different spectral zoom levels. In: *Optical coherence imaging techniques and imaging in scattering media II*, Munich, SPIE Proceedings. Optical Society of America, p 104160L. <https://doi.org/10.1117/12.2287484>.
- Prykäri T, Czajkowski J, Alarousu E, Myllylä R (2010) Optical coherence tomography as an accurate inspection and quality evaluation technique in paper industry. *Opt Rev* 17(3): 218–222. <https://doi.org/10.1007/s10043-010-0039-y>
- Qi B, Himmer AP, Gordon LM, Yang XDV, Dickensheets LD, Vitkin IA (2004) Dynamic focus control in high-speed optical coherence tomography based on a microelectromechanical mirror. *Opt Commun* 232(1–6):123–128
- Rodriguez A, Kang JU, Huang Y (2011) Real-time 4D non-invasive subsurface corrosion inspection using ultrahigh-speed, Fourier-domain optical coherence tomography. In: *International society for optics and photonics*, p 81550W
- Sabuncu M, Akdoğan M (2014) Utilizing optical coherence tomography in the nondestructive and noncontact measurement of egg shell thickness. *Sci World J* 4. <https://doi.org/10.1155/2014/205191>
- Sabuncu M, Akdoğan M (2015) Photonic imaging with optical coherence tomography for quality monitoring in the poultry industry: a preliminary study. *Rev Bras Ciênc Avícola* 17:319–324

- Sabuncu M, Ozdemir H, Akdogan MU (2017) Automatic identification of weave patterns of checked and colored fabrics using optical coherence tomography. *IEEE Photonics J* 9(5):1–8. <https://doi.org/10.1109/JPHOT.2017.2742148>
- Sandrian MG, Tomlins PH, Woolliams PD, Rasakanthan J, Lee GCB, Yang A, Považay B, Alex A, Sugden K, Drexler W (2012) Three-dimensional calibration targets for optical coherence tomography. In: Nordstrom RJ, Coté GL (eds) *Optical diagnostics and sensing XII*. SPIE proceedings, vol 8229. <https://doi.org/10.1117/12.907748>
- Schmitt R, Ackermann P (2016) OCT for process monitoring of laser transmission welding. *Laser Tech J* 13(5):15–18. <https://doi.org/10.1002/laj.201600030>
- Schmitt JM, Lee SL, Yung KM (1997) An optical coherence microscope with enhanced resolving power in thick tissue. *Opt Commun* 142(4–6):203–207. [https://doi.org/10.1016/S0030-4018\(97\)00280-0](https://doi.org/10.1016/S0030-4018(97)00280-0)
- Schmitt R, Mallmann G, Winands K, Pothen M (2012) Inline process metrology system for the control of laser surface structuring processes. *Phys Procedia* 39:814–822. <https://doi.org/10.1016/j.phpro.2012.10.105>
- Schneider S, Eppler F, Weber M, Olowojoba G, Weiss P, Hübner C, Mikonsaari I, Freude W, Koos C (2016) Multiscale dispersion-state characterization of nanocomposites using optical coherence tomography. *Sci Rep* 6:31733. <https://doi.org/10.1038/srep31733>
- Seck HL, Zhang Y, Soh YC (2011) Optical coherence tomography by using frequency measurements in wavelength domain. *Opt Express* 19(2):1324–1334. <https://doi.org/10.1364/OE.19.001324>
- Shen Y, Huang C, Monroy GL, Janjaroen D, Derlon N, Lin J, Espinosa-Marzal R, Morgenroth E, Boppert SA, Ashbolt NJ, Liu W-T, Nguyen TH (2016) Response of simulated drinking water biofilm mechanical and structural properties to long-term disinfectant exposure. *Environ Sci Technol* 50(4):1779–1787. <https://doi.org/10.1021/acs.est.5b04653>
- Shirazi M, Jeon M, Kim J (2017) Structural analysis of polymer composites using spectral domain optical coherence tomography. *Sensors* 17(5):1155
- Shu X, Beckmann L, Zhang HF (2017) Visible-light optical coherence tomography: a review. *J Biomed Opt* 22(12):1–14. <https://doi.org/10.1117/1.Jbo.22.12.121707>
- Song S, Xu J, Wang RK (2016) Long-range and wide field of view optical coherence tomography for in vivo 3D imaging of large volume object based on akinetic programmable swept source. *Biomed Opt Express* 7(11):4734–4748. <https://doi.org/10.1364/BOE.7.004734>
- Spöller F, Kray S, Grychtol P, Hermes B, Bornemann J, Först M, Kurz H (2007) Simultaneous dual-band ultra-high resolution optical coherence tomography. *Opt Express* 15(17):10832–10841
- Spring M, Liang H, Peric B, Saunders D, Podoleanu A (2008) Optical coherence tomography—a tool for high resolution non-invasive 3D-imaging of the subsurface structure of paintings. In: International Council of Museums Conservation Committee Triennial Meeting ICOM-CC, New Delhi, 22–26 September 2008, vol. II, p. 633–640. <http://irep.ntu.ac.uk/id/eprint/14553>
- Stifter D, Burgholzer P, Höglinger O, Götzinger E, Hitzenberger CK (2003) Polarisation-sensitive optical coherence tomography for material characterisation and strain-field mapping. *Appl Phys A* 76(6):947–951. <https://doi.org/10.1007/s00339-002-2065-5>
- Stifter D, Wiesauer K, Wurm M, Schlottauer E, Kastner J, Pircher M, Götzinger E, Hitzenberger CK (2008) Investigation of polymer and polymer/fibre composite materials with optical coherence tomography. *Meas Sci Technol* 19(7):074011. <https://doi.org/10.1088/0957-0233/19/7/074011>
- Stifter D, Leiss-Holzinger E, Major Z, Baumann B, Pircher M, Götzinger E, Hitzenberger CK, Heise B (2010) Dynamic optical studies in materials testing with spectral-domain polarization-sensitive optical coherence tomography. *Opt Express* 18(25):25712–25725. <https://doi.org/10.1364/OE.18.025712>
- Strupler M, Beckley AM, Benbouja F, Dubois S, Noiseux I, Mermut O, Bouchard J-P, Boudoux C (2015) Toward an automated method for optical coherence tomography characterization. *J Biomed Optics*, 20(12), 126007. <https://doi.org/10.1117/1.JBO.20.12.126007>
- Sun H, Gang T, Hu M, Liu N, Tong R, Liu X, Zhang Q, Zhou J (2016) 1310 nm source spectral-domain optical coherence tomography for Chinese pigment and jadeite research. *Optics Photonics J* 06(08):5. <https://doi.org/10.4236/opj.2016.68B009>

- Szklumowski M, Wojtkowski M, Bajraszewski T, Gorczynska I, Targowski P, Wasilewski W, Kowalczyk A, Radzewicz C (2005) Quality improvement for high resolution in vivo images by spectral domain optical coherence tomography with supercontinuum source. *Opt Commun* 246(4–6):569–578. <https://doi.org/10.1016/j.optcom.2004.11.024>
- Targowski P, Gora M, Bajraszewski T, Szklumowski M, Rouba B, Łekawa-Wysłouch T, Tymińska-Widmer L (2006a) Optical coherence tomography for tracking canvas deformation. *Laser Chem* 2006:93658. <https://doi.org/10.1155/2006/93658>
- Targowski P, Gora M, Wojtkowski M (2006b) Optical coherence tomography for artwork diagnostics. *Laser Chem* 2006:35373. <https://doi.org/10.1155/2006/35373>
- Thrane L, Jørgensen TM, Jørgensen M, Krebs FC (2012) Application of optical coherence tomography (OCT) as a 3-dimensional imaging technique for roll-to-roll coated polymer solar cells. *Sol Energy Mater Sol Cells* 97:181–185. <https://doi.org/10.1016/j.solmat.2011.10.004>
- Trinh TA, Li W, Han Q, Liu X, Fane AG, Chew JW (2018) Analyzing external and internal membrane fouling by oil emulsions via 3D optical coherence tomography. *J Membr Sci* 548:632–640. <https://doi.org/10.1016/j.memsci.2017.10.043>
- Tripathi R, Nassif N, Nelson JS, Park BH, de Boer JF (2002) Spectral shaping for non-Gaussian source spectra in optical coherence tomography. *Opt Lett* 27(6):406–408
- Tsai M-T, Chang F-Y, Yao Y-C, Mei J, Lee Y-J (2015) Optical inspection of solar cells using phase-sensitive optical coherence tomography. *Sol Energy Mater Sol Cells* 136:193–199. <https://doi.org/10.1016/j.solmat.2015.01.016>
- Vakoc B, Yun S, de Boer J, Tearney G, Bouma B (2005) Phase-resolved optical frequency domain imaging. *Opt Express* 13(14):5483–5493
- Van der Jeugt S, Bradu A, Podoleanu AG (2010) Real-time resampling in Fourier domain optical coherence tomography using a graphics processing unit. *J Biomed Opt* 15(3):030511
- Verboven P, Nemeth A, Abera MK, Bongaers E, Daelemans D, Estrade P, Herremans E, Hertog M, Saeys W, Vansteels E, Verlinden B, Leitner M, Nicolaï B (2013) Optical coherence tomography visualizes microstructure of apple peel. *Postharvest Biol Technol* 78:123–132. <https://doi.org/10.1016/j.postharvbio.2012.12.020>
- Walther J, Koch E (2014) Relation of joint spectral and time domain optical coherence tomography (jSTdOCT) and phase-resolved Doppler OCT. *Opt Express* 22(19):23129–23146
- Walther J, Gaertner M, Cimalla P, Burkhardt A, Kirsten L, Meissner S, Koch E (2011) Optical coherence tomography in biomedical research. *Anal Bioanal Chem* 400(9):2721–2743
- Wang XJ, Milner TE, Nelson JS (1995) Characterization of fluid flow velocity by optical Doppler tomography. *Opt Lett* 20(11):1337–1339
- Wang TS, Pfeiffer T, Regar E, Wieser W, van Beusekom H, Lancee CT, Springeling G, Krabbendam I, van der Steen AFW, Huber R, van Soest G (2015) Heartbeat OCT: in vivo intravascular megahertz-optical coherence tomography. *Biomed Opt Express* 6(12):5021–5032
- Webster PJL, Muller MS, Fraser JM (2007) High speed in situ depth profiling of ultrafast micromachining. *Opt Express* 15(23):14967–14972. <https://doi.org/10.1364/OE.15.014967>
- Wehrmeister U, Goetz H, Jacob DE, Soldati A, Xu W, Duschner H, Hofmeister W (2008) Visualization of the internal structure of freshwater cultured pearls by computerized x-ray microtomography. *J Gemmol* 31(1/2):15–21 The Gemmological Association of Great Britain, mq-iris:0000131633
- Weiss N, Obied KETE, Kalkman J, Lammertink RGH, van Leeuwen TG (2016) Measurement of biofilm growth and local hydrodynamics using optical coherence tomography. *Biomed Opt Express* 7(9):3508–3518. <https://doi.org/10.1364/BOE.7.003508>
- West Š, Wagner M, Engelke C, Horn H (2016) Optical coherence tomography for the in situ three-dimensional visualization and quantification of feed spacer channel fouling in reverse osmosis membrane modules. *J Membr Sci* 498:345–352. <https://doi.org/10.1016/j.memsci.2015.09.047>
- Wiesauer K, Dufau ADS, Götzinger E, Pircher M, Hitzenberger CK, Stifter D (2005) Non-destructive quantification of internal stress in polymer materials by polarisation sensitive optical coherence tomography. *Acta Mater* 53(9):2785–2791. <https://doi.org/10.1016/j.actamat.2005.02.034>
- Wiesauer K, Pircher M, Götzinger E, Hitzenberger CK, Engelke R, Ahrens G, Grützner G, Stifter D (2006) Transversal ultrahigh-resolution polarization-sensitive optical coherence tomography for strain mapping in materials. *Opt Express* 14(13):5945–5953. <https://doi.org/10.1364/OE.14.005945>

- Wiesauer K, Pircher M, Götzinger E, Hitzenberger CK, Oster R, Stifter D (2007) Investigation of glass-fibre reinforced polymers by polarisation-sensitive, ultra-high resolution optical coherence tomography: internal structures, defects and stress. *Compos Sci Technol* 67(15): 3051–3058. <https://doi.org/10.1016/j.compscitech.2007.04.018>
- Wieser W, Biedermann BR, Klein T, Eigenwillig CM, Huber R (2010) Multi-Megahertz OCT: high quality 3D imaging at 20 million A-scans and 4.5 Gvoxels per second. *Opt Express* 18(14): 14685–14704. <https://doi.org/10.1364/OE.18.014685>
- Wijesinghe REH, Lee S-Y, Jung H-Y, Jeon M, Kim J (2017a) The validity of the diagnosis of plant leaf infections using non-destructive optical inspection technique and depth-scan signal analysis. *Int J Appl Eng Res* 12
- Wijesinghe REH, Lee S-Y, Kim P, Jung H-Y, Jeon M, Kim J (2017b) Optical sensing method to analyze germination rate of Capsicum annum seeds treated with growth-promoting chemical compounds using optical coherence tomography. *J Biomed Optics*, 22(9), 091502. <https://doi.org/10.1117/1.JBO.22.9.091502>
- Yang VXD, Gordon M, Sj T, Marcon N, Gardiner G, Qi B, Bisland S, Seng-Yue E, Lo S, Pekar J, Wilson B, Vitkin I (2003) High speed, wide velocity dynamic range Doppler optical coherence tomography (part III): in vivo endoscopic imaging of blood flow in the rat and human gastrointestinal tracts. *Opt Express* 11(19):2416–2424
- Yao J, Meemon P, Lee K-S, Rolland JP (2013) Nondestructive metrology by optical coherence tomography empowering manufacturing iterations of layered polymeric optical materials. *Optical Engineering*, 52(11), 112111
- Yao J, Meemon P, Ponting M, Rolland JP (2015a) Angular scan optical coherence tomography imaging and metrology of spherical gradient refractive index preforms. *Opt Express* 23(5): 6428–6443. <https://doi.org/10.1364/OE.23.006428>
- Yao J, Xu D, Zhao N, Rolland JP (2015b) Freeform metrology using swept-source optical coherence tomography with custom pupil-relay precision scanning configuration. *Optifab* 2015. Vol. 9633. International Society for Optics and Photonics, 2015
- Yao J, Thompson KP, Ma B, Ponting M, Rolland JP (2016) Volumetric rendering and metrology of spherical gradient refractive index lens imaged by angular scan optical coherence tomography system. *Opt Express* 24(17):19388–19404. <https://doi.org/10.1364/OE.24.019388>
- Yao J, Anderson A, Rolland JP (2018) Point-cloud noncontact metrology of freeform optical surfaces. *Opt Express* 26(8):10242–10265. <https://doi.org/10.1364/OE.26.010242>
- Yasuno Y, Hong Y, Makita S, Yamanari M, Akiba M, Miura M, Yatagai T (2007) In vivo high-contrast imaging of deep posterior eye by 1-μm swept source optical coherence tomography and scattering optical coherence angiography. *Opt Express* 15(10):6121–6139
- Yi LY, Sun LQ, Ding WW (2017) Multifocal spectral-domain optical coherence tomography based on Bessel beam for extended imaging depth. *J Biomed Opt* 22(10):1–8. <https://doi.org/10.1117/1.Jbo.22.10.106016>
- Yin BW, Hyun C, Gardecki JA, Tearney GJ (2017) Extended depth of focus for coherence-based cellular imaging. *Optica* 4(8):959–965. <https://doi.org/10.1364/Optica.4.000959>
- Yun SH, Tearney G, de Boer J, Bouma B (2004) Motion artifacts in optical coherence tomography with frequency-domain ranging. *Opt Express* 12(13):2977–2998
- Zeng N, He Y, Ma H (2008) Application of optical coherence tomography in nacre identification and characterization. *Microw Opt Technol Lett* 50(2):442–445
- Zhang Y, Dong B, Bai Y, Ye S, Lei Z, Zhou Y (2015) Measurement of depth-resolved thermal deformation distribution using phase-contrast spectral optical coherence tomography. *Opt Express* 23(21):28067–28075. <https://doi.org/10.1364/OE.23.028067>
- Zhou Y, Liu T, Shi Y, Chen Z, Mao J, Zhou W (2016) Automated internal classification of beadless Chinese ZhuJi freshwater pearls based on optical coherence tomography images. *Sci Rep* 6:33819. <https://doi.org/10.1038/srep33819>
- Zhou Y, Wu D, Hui G, Mao J, Liu T, Zhou W, Zhao Y, Chen Z, Chen F (2018) Loquat bruise detection using optical coherence tomography based on microstructural parameters. *Food Anal Methods*. <https://doi.org/10.1007/s12161-018-1246-6>



# Ellipsometry

16

Jian Chen and Qiwen Zhan

## Contents

Introduction .....	514
Physical/Technical Background .....	515
Imaging Microellipsometer .....	517
Working Principle .....	517
Instrumentation .....	520
Rotationally Symmetric Scanning Microellipsometer .....	521
Microellipsometer with Rotational Symmetry .....	521
Nulling Microellipsometer with Rotational Symmetry .....	525
Applications of Microellipsometers .....	529
Measurement of Resolved Features .....	529
Measurement of Surface Features Beyond the Diffraction Limit .....	531
Measurement of Axial Birefringence .....	535
Measurement of Micro-optical Components .....	536
Rapid Mueller Matrix Polarimetry .....	537
Summary .....	539
References .....	539

---

J. Chen

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China  
e-mail: [chenjian6789@163.com](mailto:chenjian6789@163.com)

Q. Zhan (✉)

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China

Department of Electro-Optics and Photonics, University of Dayton, Dayton, OH, USA  
e-mail: [qzhan1@udayton.edu](mailto:qzhan1@udayton.edu)

**Abstract**

Over the past few decades, ellipsometry has become an important and widely used optical metrology tool in scientific research and industrial manufacturing due to its nondestructive nature and inherent simplicity. There have been increasing needs to extend the capabilities of ellipsometry to spatially resolve materials and structures with micro- and nanoscale features. In this chapter, the basic principle of ellipsometry will be reviewed, and two microellipsometry techniques that employ high numerical aperture (NA) objective lens capable of characterizing structures and materials with high spatial resolution are presented. The use of high NA lens collects more oblique rays to increase the ellipsometric signal while providing high spatial resolution of these techniques for spatially resolved characterization tasks. The working principles and experimental setups in each case are discussed mathematically and demonstrated schematically, respectively. Applications of these techniques to characterize patterned surface structures, unresolved surface features, residual axial birefringence, and micro-optical components are subsequently summarized. Finally, taking advantage of the recent rapid developments in vector optical fields, rapid polarimetric measurement techniques utilizing parallel polarization measurement that can extract the entire polarization information from the sample with a single intensity snapshot will be described.

---

**Introduction**

Optical inspection and metrology are commonly applied to materials diagnostics and characterization. From the simple qualitative information obtained from imaging and phase-contrast optical microscopy to the high-precision measurements provided by interferometry, polarimetry, and ellipsometry, optical techniques have been instrumental in the advancement of materials science, microelectronics, biology, and many other disciplines (Herman 1996). Due to the nondestructive nature and their inherent simplicity, optical measurements play an important role for both modern basic research and industrial manufacturing.

Thin film structures are widely involved in the biological sciences, materials research, optics and photonics, data storage, semiconductor, and microelectronics industries. It is critical for modern fabrication to accurately characterize the properties of these thin films. Because of the great advances in nanolithography and micro-fabrication, the ever-decreasing feature sizes in industrial fabrication have severely stretched the capabilities of optical characterization techniques and put a strong requirement on characterizing patterned structures. As the vector nature of optical field, optical polarization offers attractive opportunities in many optical measurement techniques. Meanwhile, the polarization diversity within the beam cross section of vectorial optical fields has attracted increasing interests in optical measurement and may provide unique solutions to some of the metrology challenges.

In this chapter, two spatially resolved ellipsometric techniques that can characterize materials with high spatial resolution are presented. Applications of these techniques to characterize patterned surface structures, unresolved surface features, residual axial birefringence and micro-optical components are subsequently presented. Then, a rapid Mueller matrix polarimetry enabled by the parallel polarization measurement with the use of vector beams that can extract 12 Mueller matrix elements from a single reflective intensity snapshot image is also described and discussed.

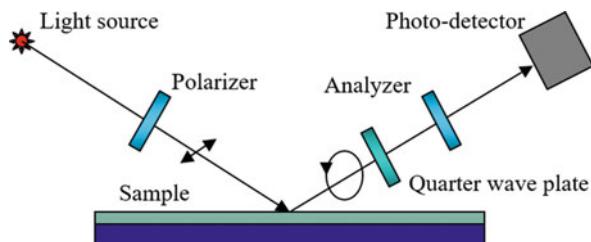
## Physical/Technical Background

Ellipsometry is one of the most powerful optical metrological techniques to measure the properties of thin films, which could simultaneously determine the film thickness and index of refraction in an accurate and nondestructive manner (Azzam and Bashara 1977; Tompkins 1993; Tompkins and McGahan 1999). Generally speaking, optical ellipsometry can be defined as the measurement of the state of polarization of the light wave. This technique is based on utilizing the polarization modification that occurs as the polarized light is transmitted through or reflected from the interfaces of thin films. The information about the interface, such as the thickness of the film, the refraction index of the material, the roughness of the interface, etc., can be extracted via measuring the modification in conjunction with an appropriate model of the structure under investigation. To illustrate the working principle, the schematic of a nulling ellipsometry setup is shown in Fig. 1, in which the surface of a sample is illuminated with monochromatic light with a known and controllable state of polarization at a fairly large angle. The polarization of the reflected light is then analyzed by adjusting a quarter-wave plate and an analyzer to realize a null signal on the detector. According to the angular positions of the wave plate and analyzer under the null condition, the fundamental quantities ( $\Psi$ ,  $\Delta$ ) pair of the ellipsometry can be defined as:

$$\frac{r_p}{r_s} = \tan \Psi \cdot e^{j\Delta}, \quad (1)$$

where the  $r_p$  and  $r_s$  are the complex reflection coefficients for TM and TE polarized light upon reflection from the interface, respectively. Then, based on the ( $\Psi$ ,  $\Delta$ ) pair,

**Fig. 1** Conceptual schematic of nulling ellipsometer



information about the sample can be deduced with proper optical modeling and numerical regression method.

Due to its amazing accuracy and sensitivity, ellipsometry has been widely employed in many thin film-related disciplines, including materials characterization and diagnostics, optics and photonics industry, semiconductor and microelectronics industry, biological and biochemical sciences, etc. However, because of the limitation of the illumination spot size, the spatial resolution of a standard ellipsometer is typically on the order of a few hundred micrometers. Therefore, the poor spatial resolution of conventional ellipsometry severely hinders its application to characterize samples with fine features. Consequently, spatially resolved ellipsometry is in great demand for the characterization and diagnostics of patterned thin films and surfaces.

Objective lenses can be adopted to improve the spatial resolution of the ellipsometer. One approach is to modify the standard ellipsometer setup by inserting an objective lens in the incident arm (Erman and Theeten 1986; Liu et al. 1994; Albersdörfer et al. 1998; Jin et al. 1996). However, in order to avoid the collision of the lens and the sample, the highest available numerical aperture (NA) is limited by the slant angle between the objective and the sample. In addition, the use of higher NA lens in this configuration will lead to the integration of ellipsometric signals over a large incident angular range, leading to reduced measurement sensitivity. Consequently, this method only provides moderate improvement of spatial resolution. Another technique exploits normal incidence and large NA objective lenses to achieve high spatial resolution (Gold et al. 1989; Leng et al. 1998; See et al. 1996). In this approach, the incident light is focused to a diffraction-limited spot and reflected by the sample surface. The change of the polarization state of light at the sample is then analyzed at the back focal plane of the objective lens. Since the fields at the back focal plane and at the sample plane are related through a Fourier transform in this technique, each incidence angle is mapped to a specific position in the back focal plane. Through analyzing the state of polarization at the back focal plane, the properties of the sample within the focus can be inferred. In order to accurately measure the ellipsometric signal for a given incident angle, the pixel size at the back focal plane should be small. However, the small pixel size of the detector required for good ellipsometric performance often also leads to a low signal-to-noise ratio (SNR) in the measurement.

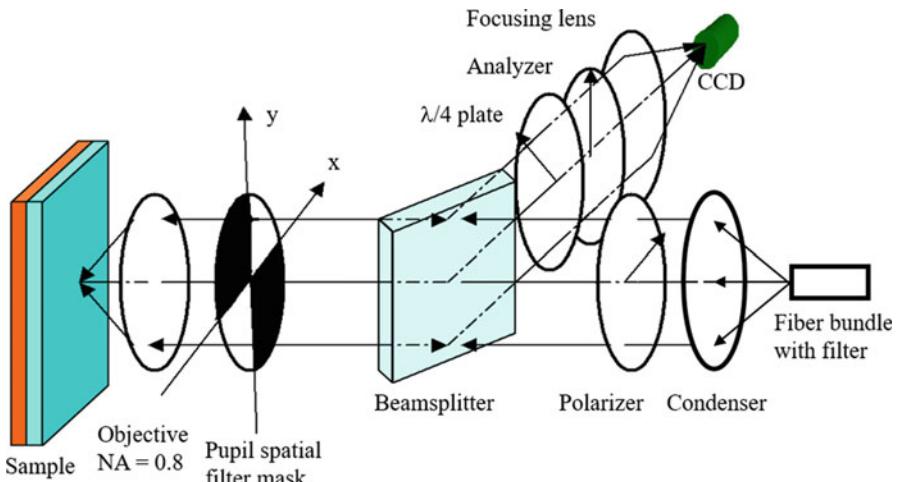
In the following, two spatially resolved ellipsometer techniques that employ high NA objective lens and normal incidence will be presented and discussed. The high NA lens contains more oblique rays to increase the ellipsometric signal. Meanwhile, the high NA serves to increase the spatial resolution of these techniques. First, an imaging microellipsometer is introduced. This imaging microellipsometer can characterize the sample with high resolution and form an image of the sample in parallel. Compared to existing techniques, this microellipsometer has a very simple setup to perform accurate measurements with higher speed and higher resolution, making the ellipsometer a very powerful tool for industrial and scientific research. In another microellipsometer design, a rotational symmetry concept is adopted to enhance the SNR of the microellipsometry. This rotationally symmetric scanning

microellipsometer can provide high spatial resolution characterization capability with high SNR and optimal ellipsometric performance.

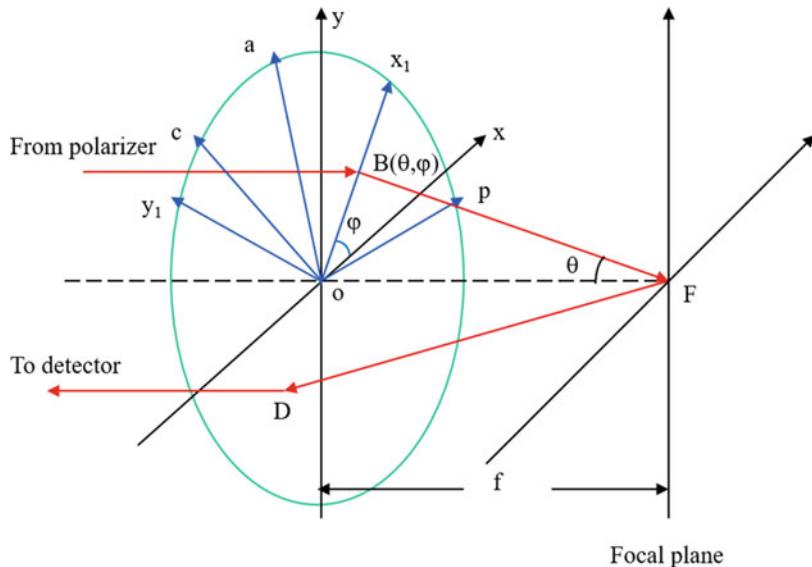
## Imaging Microellipsometer

### Working Principle

The imaging microellipsometer is schematically shown in Fig. 2. A high numerical aperture objective is adopted in the setup. The transmission axes of the polarizer and the analyzer are aligned with the  $x$ -axis and  $y$ -axis, respectively. The sample is placed at the focal plane of the objective lens. A fiber bundle with a narrow band filter serves as the illumination source to form an image on the CCD. This can be replaced with any other extended optical illumination source. The condenser collimates the light from each point of the source. Since each point of the source in the field of view of the condenser subtends very small angles with respect to the optical axis, each image point can be analyzed in the same way as the on-axis point. A geometric optics model is shown in Fig. 3 to give an intuitive description of the setup;  $op$  denotes the transmission axis of the polarizer,  $oa$  denotes the transmission axis of the analyzer, and  $oc$  is the  $c$ -axis of the  $\lambda/4$  plate. For a ray from the on-axis point entering the pupil at position  $(\theta, \varphi)$ , where  $\theta$  is the corresponding incident angle of this ray on the sample after the lens and  $\varphi$  is the azimuth angular position relative to the  $x$ -axis, the power of this ray on the detector is (Zhan and Leger 2002a):



**Fig. 2** Diagram of the imaging microellipsometer design



**Fig. 3** Coordinates adopted in the geometric optics model

$$\begin{aligned}
 P(\theta, \varphi) &= \left| \vec{E}(\theta, \varphi) \right|^2 dS \\
 &= \left[ \frac{1}{2} |r_p|^2 \cos^2 \varphi + \frac{1}{2} |r_s|^2 \sin^2 \varphi - \frac{1}{2} |r_p| |r_s| \sin 2\varphi \sin \delta_{ps} \right] dS, \quad (2)
 \end{aligned}$$

when  $C = \pi/4$  and

$$\begin{aligned}
 P(\theta, \varphi) &= \left| \vec{E}(\theta, \varphi) \right|^2 dS \\
 &= \left[ \frac{1}{2} |r_p|^2 \cos^2 \varphi + \frac{1}{2} |r_s|^2 \sin^2 \varphi + \frac{1}{2} |r_p| |r_s| \sin 2\varphi \sin \delta_{ps} \right] dS, \quad (3)
 \end{aligned}$$

when  $C = -\pi/4$ .

From Eqs. (2) and (3), it is clear that the phase delay between the TE and TM components is in the interference term (the last term in each expression). Meanwhile, if the entire aperture is used, the integral of  $\sin 2\varphi$  over  $(0, 2\pi)$  will eliminate the phase information. Thus, a spatial filter is required to break the azimuth symmetry and extract the phase delay. One simple way is to use an aperture to block two opposite quadrants of the lens pupil. With this mask, two power signals can be obtained by integrating Eqs. (2) and (3) over the aperture. For  $C = \pi/4$ ,

$$\begin{aligned}
I_1 &= K^2 \int_0^{\theta_{max}} f^2 \sin \theta d\theta \int_{\Phi} d\varphi \left[ \frac{1}{2} |r_p|^2 \cos^2 \varphi + \frac{1}{2} |r_s|^2 \sin^2 \varphi - \frac{1}{2} |r_p| |r_s| \sin 2\varphi \sin \delta_{ps} \right] \\
&= \frac{K^2}{2} \int_0^{\theta_{max}} f^2 \sin \theta d\theta \int_{\Phi} d\varphi \left[ |r_p|^2 \cos^2 \varphi + |r_s|^2 \sin^2 \varphi \right] \\
&\quad - \frac{K^2}{2} \int_0^{\theta_{max}} f^2 \sin \theta d\theta \int_{\Phi} d\varphi [|r_p| |r_s| \sin 2\varphi \sin \delta_{ps}] \\
&= I_0 - I_{int}.
\end{aligned} \tag{4}$$

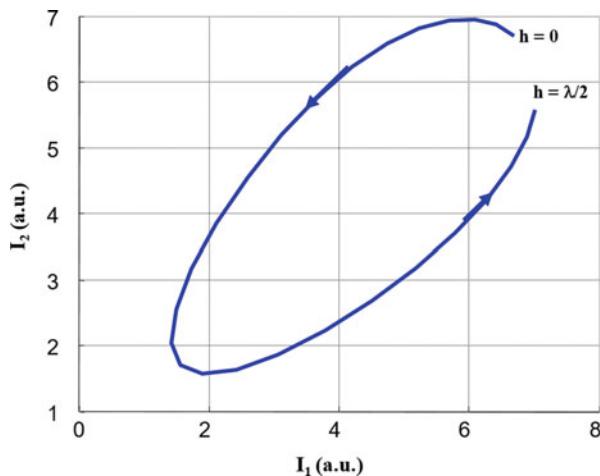
For  $C = -\pi/4$ ,

$$\begin{aligned}
I_2 &= K^2 \int_0^{\theta_{max}} f^2 \sin \theta d\theta \int_{\Phi} d\varphi \left[ \frac{1}{2} |r_p|^2 \cos^2 \varphi + \frac{1}{2} |r_s|^2 \sin^2 \varphi + \frac{1}{2} |r_p| |r_s| \sin 2\varphi \sin \delta_{ps} \right] \\
&= \frac{K^2}{2} \int_0^{\theta_{max}} f^2 \sin \theta d\theta \int_{\Phi} d\varphi \left[ |r_p|^2 \cos^2 \varphi + |r_s|^2 \sin^2 \varphi \right] \\
&\quad + \frac{K^2}{2} \int_0^{\theta_{max}} f^2 \sin \theta d\theta \int_{\Phi} d\varphi [|r_p| |r_s| \sin 2\varphi \sin \delta_{ps}] \\
&= I_0 + I_{int},
\end{aligned} \tag{5}$$

where  $K$  is a constant,  $\Phi$  denotes the azimuth integration interval of  $[0, \pi/2] \cup [\pi, 3\pi/2]$ ,  $\theta_{max}$  corresponds to the NA, and  $I_{int}$  is the interference term. The  $I_1$  and  $I_2$  will be used as the ellipsometric parameters in the measurement, instead of the conventional  $(\Psi, \Delta)$  pair. This pupil plane spatial filtering technique is equivalent to a quad detection adopted in some back focal plane ellipsometry techniques (Leng et al. 1998). However, this pupil plane spatial filtering can be used to perform ellipsometric measurements in parallel to generate a two-dimensional image without scanning. The nonuniformity of the light source in the transverse dimension needs to be taken into consideration in the imaging mode. The non-uniformity could be calibrated by a standard sample such as a mirror. However, a much simpler approach was adopted by taking the ratio between  $I_1$  and  $I_2$  as the signal to eliminate this nonuniformity arising from the extended illumination source.

A more rigorous Richards-Wolf vectorial diffraction model (Wolf 1959; Richards and Wolf 1959) including the diffraction effect from high NA objective lens has also been developed for the simulation of the optical system. According to the simulation results, a spatial filter at the back focal plane is necessary for extracting information about the sample. A filter (aperture mask) is used to block the II and IV quadrants of

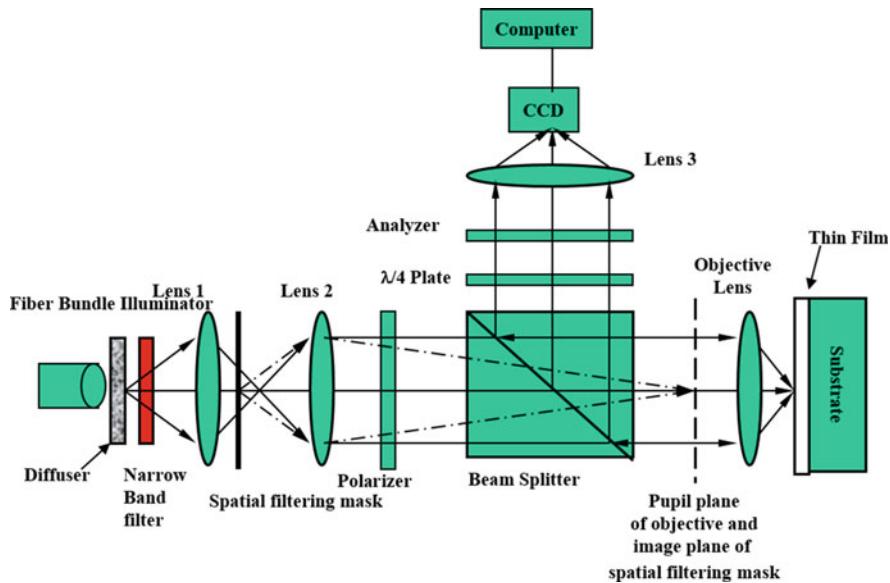
**Fig. 4** Simulation results for a thin film sample based on Richards-Wolf theory



the objective lens' exit pupil in the simulation, and the two intensities corresponding to two orthogonal angular positions of the  $\lambda/4$  plate are calculated. Figure 4 presents the two ellipsometric intensity signals as a function of film thickness ( $h$ ) for a thin SiO<sub>2</sub> film on a silicon substrate. The simulation results clearly show the potential of using the imaging microellipsometer setup to characterize samples with high spatial resolution.

## Instrumentation

Figure 5 gives the actual experimental setup of the imaging ellipsometer, in which a fiber bundle illuminator is used as an extended source to form an image of the object and permit analysis of each sample point in parallel. A diffuser is placed against the fiber bundle to homogenize the light. And a 10 nm bandwidth interference bandpass filter with central wavelength of 633 nm is placed after the diffuser. Lens1 and lens2 are employed to collimate each point of the source and project the quadrant spatial filter to the pupil plane of the objective. The linearly polarized incident light is focused onto the sample through an objective ( $50\times/\text{NA} = 0.8$ ) with 4 mm focal length. After reflection from the sample, the light is collected by the objective and steered to the analyzer arm by the beamsplitter. Generally speaking, the reflectivity of the beamsplitter for the  $s$  and  $p$  component is different, which introduces an extra ellipsometric rotation to the signal. Hence, the difference has to be included in the model in order to generate an accurate lookup table. For the used beamsplitter,  $r_p/r_s = 0.6378 - 0.1125j$  is determined experimentally and included in the optical model used. A zero-order polymer  $\lambda/4$  plate with  $+/-9^\circ$  angular acceptance is located in the analyzer arm. An analyzer is placed after the  $\lambda/4$  plate and aligned at a crossed position with respect to the polarizer. Lens3 is a 40 cm doublet that images the object onto the CCD camera with magnification of 100. The pixel size



**Fig. 5** Schematic of the experimental setup for the imaging microellipsometer

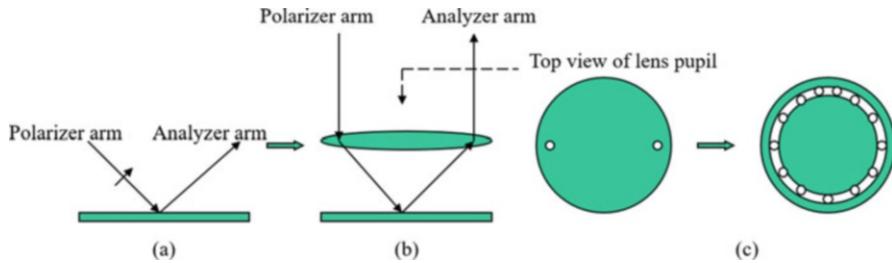
of the camera is  $9.8 \mu\text{m}$ , which corresponds to about  $0.1 \mu\text{m}$  on the object plane. The resolution of the system is  $1.22\lambda/2NA \approx 0.48 \mu\text{m}$ ; thus each resolution spot is over-sampled by a factor of about five. A 10-bit A/D converter is used as an interface between the camera and computer. The image size is  $512 \times 480$ , approximately corresponding to an area of  $51.2 \mu\text{m}$  by  $48 \mu\text{m}$  on the sample. To reduce the temporal noise, 64 frames are averaged. Two images for the two orthogonal orientations of the  $\lambda/4$  plate are captured to normalize the nonuniformity of the source by calculating their ratio. Finally, the ratio signal of each point is converted into thickness by exploiting a theoretical curve generated according to the aforementioned optical models.

## Rotationally Symmetric Scanning Microellipsometer

### Microellipsometer with Rotational Symmetry

#### Working Principle

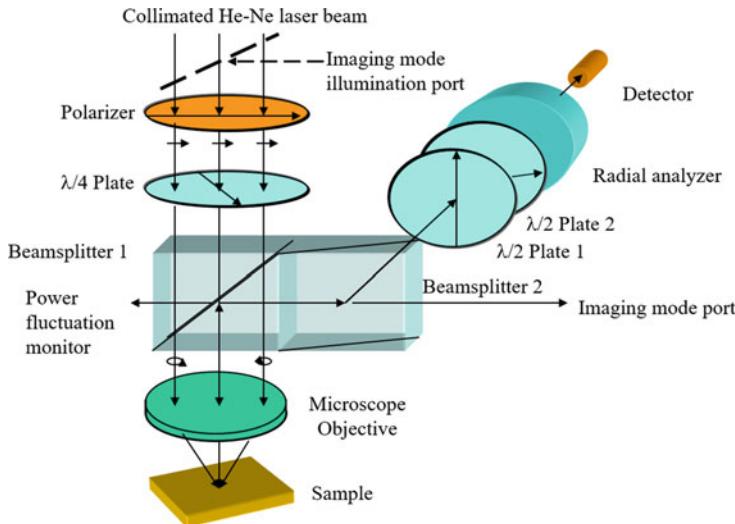
In the imaging microellipsometer technique described above, integration over a cone of incident angles is adopted to achieve high signal level at the cost of compromised ellipsometric sensitivity. The accuracy of the system will also be affected by the illumination nonuniformity. These disadvantages can be overcome by a microellipsometer based on a rotational symmetry concept, which is best illustrated in Fig. 6. A standard ellipsometer is given in Fig. 6a. In Fig. 6b, a high NA objective



**Fig. 6** Illustration of the design principle of rotationally symmetric microellipsometer. (a) Standard ellipsometer. (b) Insertion of lens. A single ray path represents a conventional ellipsometer “channel.” (c) A microellipsometer by continuously repeating the single “channel” of (b) in a rotationally symmetric manner

lens is inserted in the optical path to focus the normally incident beam to a tight spot on the sample. Each ray path in Fig. 6b acts just like a standard ellipsometer. If the rotating analyzer technique is used to detect this channel, a sinusoidal signal will be observed on the detector, and the ellipsometric information carried by this ray will also be obtained (Azzam and Bashara 1977). Considering one ray as an ellipsometric “channel” of the standard rotating analyzer, this channel can be repeated in a rotationally symmetric manner such that the microellipsometer acts like a multi-channel conventional ellipsometer (Fig. 6c). Each individual channel located at a different angular location within a common annular region will be designed to provide an identical sinusoidal signal at the detector. The high numerical aperture illumination allows ellipsometric measurements of the sample with high spatial resolution. Owing to the symmetry, each channel contributes to the signal collaboratively, and a single large-area detector can be employed to measure the total power from all the channels for subsequent analysis. As a result, the SNR of the system is improved.

To implement this idea, the rotational symmetry has to be maintained throughout the optical system. Several techniques are used in the design of such a rotationally symmetric microellipsometer. As shown in a conceptual setup illustrated in Fig. 7, a collimated laser beam goes through a polarizer, a quarter-wave plate, and beamsplitter 1. The polarizer and quarter-wave plate are adjusted to generate the circularly polarized beam illuminating the objective pupil. Because of this circular state of polarization, the illumination of each channel is equivalent except for a geometric projection phase difference (Berry’s phase) (Zhan and Leger 2002b, c). Since the detector will be utilized to detect the power, this phase difference has no significance in this microellipsometer design. The high NA objective lens tightly focuses the circularly polarized beam to a small spot on the sample and collects the reflected light. The polarization state of each channel is modified to elliptical polarization by the Fresnel reflection at the sample interface. Up to this point, the rotational symmetry has been maintained. Two identical beamsplitters are adopted to direct the reflected light to the analyzer arm. The two beamsplitters are aligned such that the s-polarization for the first beamsplitter becomes the p-polarization for the second one and vice versa. In this way, the state of polarization and the rotational



**Fig. 7** Diagram of a rotationally symmetric microellipsometer

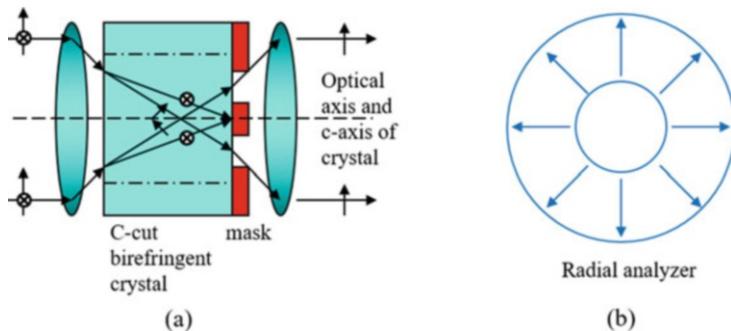
symmetry of the reflected beam are maintained. If only one beamsplitter is used, the difference of the reflectivity for the s and p components will introduce distortion in the state of polarization and consequently destroy the symmetry.

In the analyzer arm, a variable circular retarder made up of two half-wave plates is employed to introduce a spatially invariant polarization rotation (Zhan and Leger 2002c, d) (Fig. 7). The Jones matrix for this device is:

$$T = \begin{pmatrix} \cos(2\Delta\phi) & -\sin(2\Delta\phi) \\ \sin(2\Delta\phi) & \cos(2\Delta\phi) \end{pmatrix} = R(-2\Delta\phi), \quad (6)$$

where  $\Delta\phi$  is the angle between the fast axes of the two half-wave plates. This is a rotation operator with the amount of rotation independent of the initial state of polarization. Hence, the rotational symmetry of the polarization is still maintained. Then, the beam passes through a radial analyzer with transmission axis pointing in the radial direction. Figure 8 shows an example of such a radial analyzer design. Through adjusting the angle between the fast axes of the two half-wave plates while rotating one of them, an identical amount of polarization rotation can be introduced into each channel simultaneously. This is equivalent to keeping the polarization of the beam static by rotating the transmission axis of each channel simultaneously with the same amount of rotation. Thus, each channel is equivalent to a rotating analyzer ellipsometer. Meanwhile, owing to the rotational symmetry, every channel provides the same sinusoidal signal for the detector. It is this collaboration of the channels that results in high SNR compared to other microellipsometer techniques.

Assuming that for each channel, the ellipse of polarization has an ellipticity angle  $\epsilon$  and elevation angle  $\theta_0$ . After a tedious derivation using Jones calculus, the signal on the detector can be given by (Zhan and Leger 2002c):



**Fig. 8** Schematic of a radial analyzer. (a) Radial analyzer design based on birefringent cube. (b) The transmission axes pattern of this radial analyzer

$$P = K \{1 + \cos 2\varepsilon \cos [2(F - \theta_0)]\} \quad (7)$$

where  $K$  is a constant and  $F$  is the amount of polarization rotation introduced by the circularly variable retarder. The ellipsometric information is in the amplitude and phase of the signal, which can be detected by employing standard synchronous detection techniques. The  $(\varepsilon, \theta_0)$  can be related to the more commonly used ellipsometric pairs via:

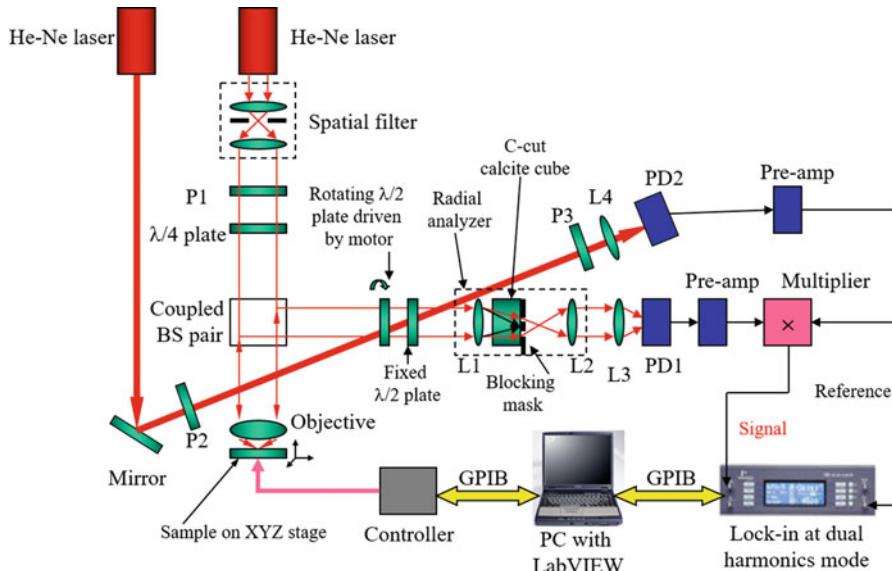
$$\tan^2 \Psi = \frac{\sin^2 \theta_0 \cos^2 \varepsilon + \cos^2 \theta_0 \sin^2 \varepsilon}{\cos^2 \theta_0 \cos^2 \varepsilon + \sin^2 \theta_0 \sin^2 \varepsilon}, \quad (8)$$

$$\Delta = \tan^{-1} \left( \frac{\tan \varepsilon}{\tan \theta_0} \right) + \tan^{-1} (\tan \theta_0 \tan \varepsilon) - \frac{\pi}{2}. \quad (9)$$

Either the  $(\varepsilon, \theta_0)$  pair or the  $(\Psi, \Delta)$  pair can be used in a standard regression procedure to infer the thickness and index of refraction of thin films within an extremely small illuminated area.

## Instrumentation

The experimental setup is demonstrated in Fig. 9. In the signal path, a high NA lens ( $\text{NA} = 0.8$ ) focuses the circularly polarized incident beam onto the sample and then collects the reflected light. The reflected light passes through the coupled beamsplitter pair, variable circular retarder, and radial analyzer and is finally focused on the photodiode PD1. A motor is used to drive the first half-wave plate at a constant speed in the variable circular retarder. The radial analyzer module consists of two lenses and a c-cut calcite cube. A preamp is used to amplify the signal. Besides the signal path, another laser was used to generate the reference signal for synchronous detection. This He-Ne laser beam passes through a polarizer, the circularly variable retarder, and another polarizer and is brought to focus on the photodiode PD2. An identical preamp is used to amplify the reference signal. In order to detect the modulation in Eq. (7), a multiplier is used to mix the signal and the reference such that



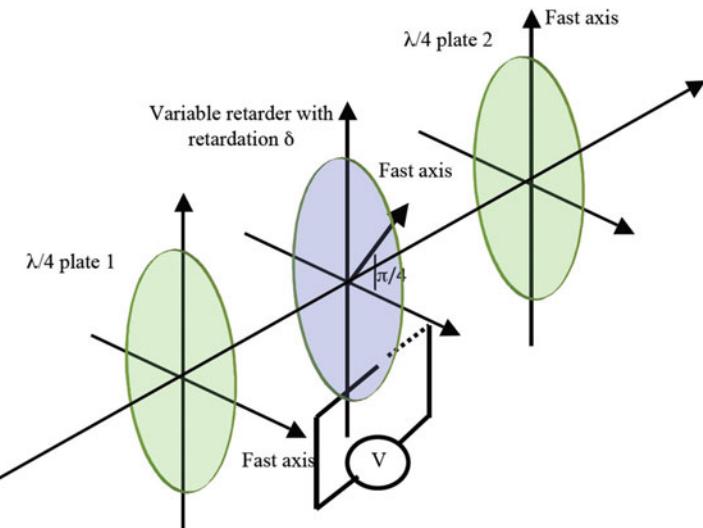
**Fig. 9** Experimental setup of the rotationally symmetric microellipsometer

the dc term is raised to the first harmonic and the ac term is raised to the second harmonic. The output of the multiplier is connected to the signal input of a lock-in amplifier running in dual-harmonics mode. The reference signal is also connected to the reference input of the lock-in amplifier. The lock-in returns both the phase and amplitude of the first and the second harmonics simultaneously. With the above measured results, one can calculate the  $(\varepsilon, \theta_0)$  pair and further derive the refractive index and thickness of the thin film through regression models used in standard ellipsometry. The sample is mounted on a translation stage with electrostrictive actuators. Ellipsometric images of the sample can be obtained through scanning. To facilitate the data acquisition, the experimental setup is fully automated. A LabVIEW program in a laptop computer is used to control the scanning, collect and process measurement data, and display the captured ellipsometric images.

## Nulling Microellipsometer with Rotational Symmetry

### Working Principle

In the microellipsometer implementation described in section “[Microellipsometer with Rotational Symmetry](#),” a mechanical motor was used to drive the polarization rotator, causing vibrational noise and limiting the modulation frequency of the polarization signal. In a continuation work, the mechanically driven polarization rotator was replaced with an electro-optics (EO) polarization rotator (as shown in Fig. 10) (Tschimwang and Zhan 2010). The use of this EO polarization rotator



**Fig. 10** Schematic of a nonmechanical polarization rotator via using electro-optics variable retarder sandwiched between two orthogonally oriented  $\lambda/4$ -plates

eliminates the vibrational noise and enables higher modulation frequency for the polarization signal. Furthermore, it makes it possible to implement a nulling detection scheme that was proposed in Zhan and Leger (2002c). The Jones matrix of these devices can be expressed as:

$$\begin{aligned} T &= R\left(-\frac{\pi}{2}\right) \begin{pmatrix} 1 & 0 \\ 0 & -j \end{pmatrix} R\left(\frac{\pi}{2}\right) R\left(-\frac{\pi}{4}\right) \begin{pmatrix} 1 & 0 \\ 0 & e^{-j\delta} \end{pmatrix} R\left(\frac{\pi}{4}\right) \begin{pmatrix} 1 & 0 \\ 0 & -j \end{pmatrix} \\ &= -je^{-j\delta/2} R\left(\frac{\delta}{2}\right), \end{aligned} \quad (10)$$

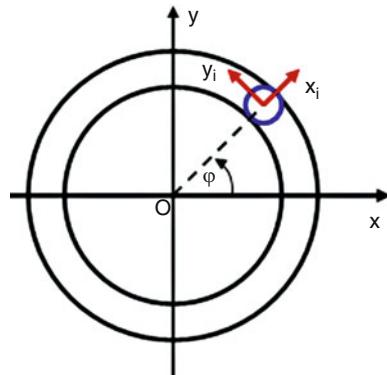
where  $\delta$  is the retardation of the EO retarder and  $R(\bullet)$  denotes the polarization rotation matrix for a given rotation angle. These devices can be made with very large clear aperture, and the amount of rotation can be adjusted and modulated.

The total signal on the detector can be derived using Jones calculus. The derivation is based on the geometric layout of the contributing channels in the pupil plane, as shown in Fig. 11. In this diagram, the illumination is a circularly polarized plane wave in the  $xOy$  coordinate system (the laboratory coordinate system), and its Jones vector is given by

$$E_{in} = \begin{pmatrix} 1 \\ j \end{pmatrix} \quad (11)$$

A ray entering the pupil at an azimuthal angle  $\varphi$  is regarded as a single channel. The coordinate system of this channel is  $x_iOy_i$  (the local coordinate system), as

**Fig. 11** Coordinate systems in the back focal plane of the objective



shown in Fig. 11. The local coordinate system is chosen such that the two eigenvectors align with the s and p directions of the corresponding incident plane wave when the ray hits the sample. In the local coordinate system, the incident polarization can be written as:

$$E_{loc} = R(\varphi)E_{in} = \begin{pmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{pmatrix} \begin{pmatrix} 1 \\ j \end{pmatrix} = e^{j\varphi} \begin{pmatrix} 1 \\ j \end{pmatrix}. \quad (12)$$

Eq. (12) shows that each local channel sees an identical circularly polarized incident light, except for a geometric phase. This phase will not affect the power of the ellipsometric signal to be detected. After reflection from the sample, and owing to the rotational symmetry, each channel carries an identical elliptical polarization signal in its own local coordinates, except for the geometric phase. Assuming that, in each local channel, there is an ellipse of polarization with an ellipticity angle  $\varepsilon$  and a major axis making an angle  $\theta_0$  with respect to the  $Ox_i$  axis. When a dc bias and an ac modulation voltage are applied across the EO modulator, the ellipse of polarization in each channel is rotated by:

$$\Phi = \Phi_{bias} + \alpha_m \cos \omega t, \quad (13)$$

where  $\Phi_{bias}$  is the part of the rotation angle attributed to the dc bias voltage,  $\alpha_m$  is the amplitude of the modulation angle attributed to the ac voltage, and  $\omega$  is the modulation frequency. The general form of the signal is similar to that of a rotating analyzer ellipsometer. In this case, the general signal can be expressed as:

$$P = K \{1 + \cos 2\varepsilon \cos [2(\Phi_{bias} + \alpha_m \cos \omega t + \theta_0)]\}, \quad (14)$$

where  $K$  is a constant. Assuming a small angular modulation, i.e.,  $\alpha_m \ll 1$ , and using the Taylor's expansion, the cosine term in Eq. (14) can be expanded as:

$$\begin{aligned} & \cos [2(\Phi_{bias} + \alpha_m \cos \omega t + \theta_0)] \\ & \cong \cos 2(\Phi_{bias} + \theta_0)(1 - \alpha_m^2 + \alpha_m^2 \cos 2\omega t) - 2\alpha_m \sin 2(\Phi_{bias} + \theta_0) \cos \omega t. \end{aligned} \quad (15)$$

Substitution of Eq. (15) into Eq. (14) reveals three most relevant harmonic components in the general signal:

$$P(0) = K[1 + \cos 2\varepsilon(1 - \alpha_m^2) \cos 2(\Phi_{\text{bias}} + \theta_0)], \quad (16)$$

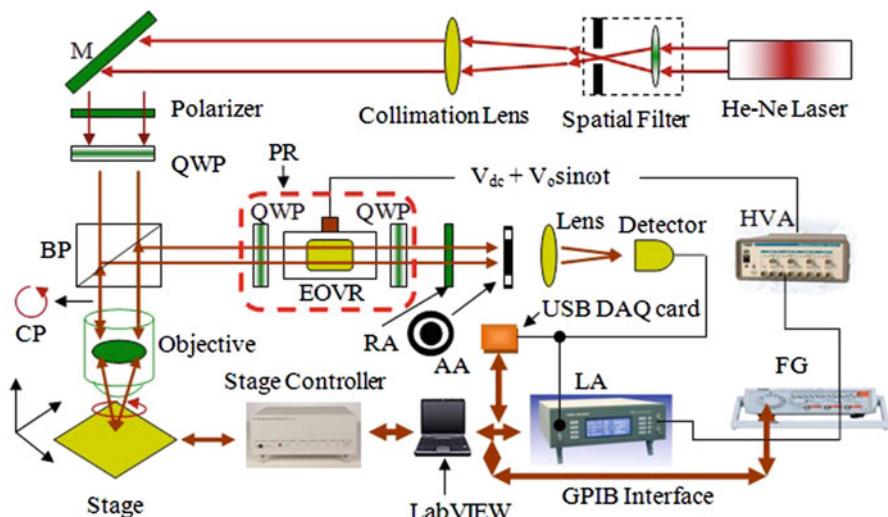
$$P(\omega) = 2K\alpha_m \cos 2\varepsilon \sin 2(\Phi_{\text{bias}} + \theta_0) \cos(\omega t + \pi), \quad (17)$$

$$P(2\omega) = K \cos 2\varepsilon \alpha_m^2 \cos 2(\Phi_{\text{bias}} + \theta_0) \cos 2\omega t. \quad (18)$$

Eqs. (16), (17), and (18) are the dc, first, and second harmonics of the signal, respectively. They all directly depend on the ellipticity angle  $\varepsilon$ , the modulation amplitude  $\alpha_m$ , and the ellipse orientation  $\theta_0$ . It can be deduced from Eq. (17) that when  $\Phi_{\text{bias}} = -\theta_0$  or  $\Phi_{\text{bias}} = -\theta_0 \pm \pi/2$ , the first harmonic becomes null, i.e.,  $P(\omega) = 0$ , hence, the nulling detection scheme. The ratio of the dc signal at two consecutive null positions, i.e.,  $\Phi_{\text{bias}} = -\theta_0$  and  $\Phi_{\text{bias}} = -\theta_0 \pm \pi/2$ , gives  $\tan^2 \varepsilon$ , which means one can determine both the elevation angle  $\Phi_{\text{bias}} = -\theta_0$  and the ellipticity angle  $\varepsilon$  by measuring two consecutive nulls.

### Instrumentation

Figure 12 presents the schematic experimental setup of this nulling microellipsometer. The beam propagates through the intermediate optical elements for spatial filtering and collimation and ends up being focused by the objective lens ( $100\times$ , NA = 0.9) onto the sample. The sample is placed on a computer-controlled XYZ translation stage, and the stage is controlled through a GPIB interface. The reflected polarization is collected by the objective lens and redirected toward the analyzer arm by a nonpolarizing beamsplitter pair, before entering the polarization



**Fig. 12** Experimental setup of a nulling microellipsometer with rotational symmetry

rotator. The electro-optics variable retarder in the polarization rotator is biased by a range of amplified voltages ( $-350\text{ V}$  to  $350\text{ V}$ ) supplied by a function generator and amplified ( $100\times$ ) by a high voltage amplifier. A small ac voltage ( $20\text{ mV}$  at  $1\text{ KHz}$ ) is supplied by the same function generator for signal modulation. The beam leaving the polarization rotator passes through a radial analyzer and then through an annular aperture for spatial filtering of the central portion of the beam attributed to low angles of illumination. After spatial filtering, the beam is focused onto the detector. The output of the detector is split into two parts: one part is sent to a lock-in amplifier for the detection of the first harmonic signal, while the other part is sent to a data acquisition card. The temporal signal acquired with the data acquisition card is preprocessed via low-pass filtering to eliminate the high frequency noise and Fourier transformed to retrieve the dc signal. The function generator, the translation stage, and the lock-in amplifier are all controlled by a LabVIEW program via GPIB interface. The data acquisition card is controlled through a USB interface.

---

## Applications of Microellipsometers

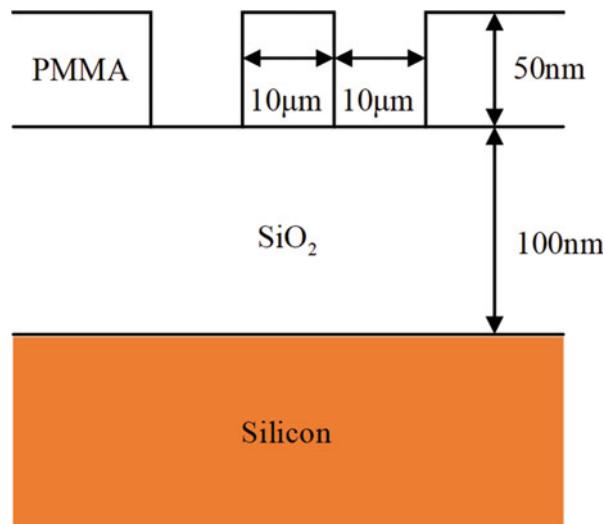
### Measurement of Resolved Features

#### Surface Topology Measurement

The capability of the imaging microellipsometer is demonstrated by applying the technique to measure a 50-nm deep PMMA (polymethyl methacrylate) grating fabricated on top of a 100 nm  $\text{SiO}_2$  thin film with a silicon wafer as a substrate (Fig. 13). Ellipsometric images were generated at two different  $\lambda/4$  plate angles ( $+\pi/4$  and  $-\pi/4$ ), and the ratio signal was displayed. The results (see Fig. 14) clearly show the existence of distinct polarization effects for different regions. Figure 14a shows the 2D imaging ellipsometric result of the PMMA grating. A cross section of the image is shown in Fig. 14b. The mesa corresponds to a structure of PMMA (50 nm)/ $\text{SiO}_2$  (100 nm)/silicon, and the trench corresponds to a structure of  $\text{SiO}_2$  (100 nm)/silicon.

From the ellipsometric signal, the thickness of each region was found by looking up the measured ratio signal in a theoretical lookup table generated by the Richards-Wolf model. From the signal, the sensitivity can be determined to be approximately  $5\text{ \AA}$ . The measured thickness is within  $\pm 5\%$  of the actual thickness. Several factors could bring errors to the absolute measurement. One possible error source is the assumed index of refraction. The imaging ellipsometer should be able to measure the thickness and refractive index using the two independent intensity measurements obtained at orthogonal  $\lambda/4$  plate angles. However, the ratio of these two intensities has been used as the analysis parameter to eliminate the effect of source nonuniformities and sample reflectance variations. From this ratio signal, only one parameter can be measured given the knowledge of the other parameter. Thus, in order to model the structure more accurately, the exact refractive index of each layer needs to be known. Other possible factors that could cause the deviation of experimental data are the residual birefringence of the lens and inaccuracies in modeling the beamsplitter.

**Fig. 13** Illustration of the PMMA grating sample

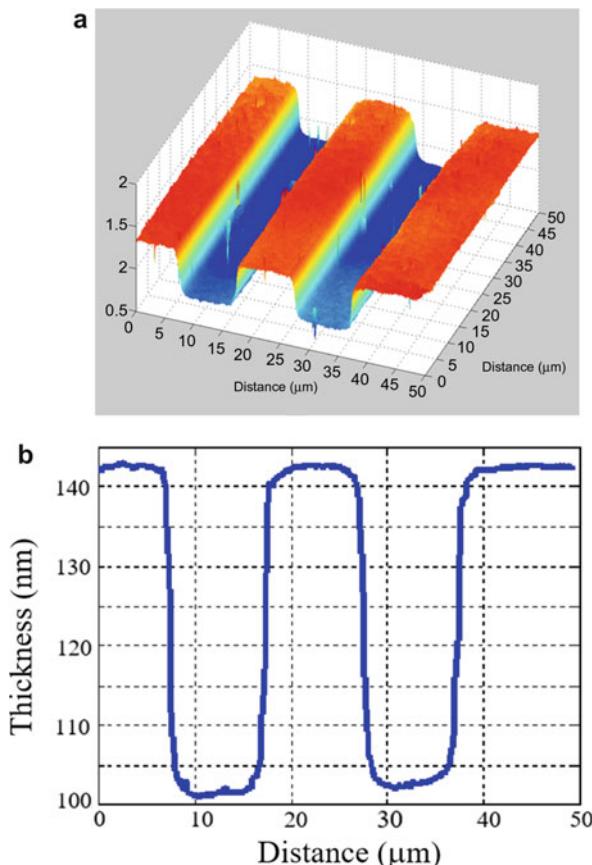


The nonuniformity of illumination at the pupil gives another possible source of error. However, a calibration procedure can be adopted to compensate the last three factors and thus improve the measurement accuracy.

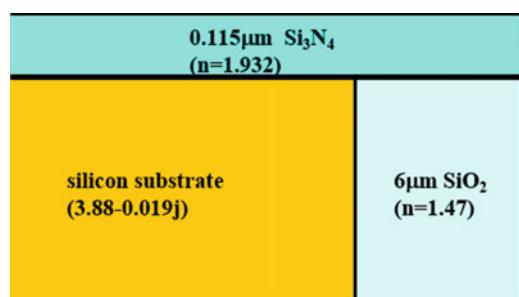
### Refractive Index Variation Measurement

A sample with index of refraction variation has also been measured with the imaging microellipsometer. The sample is prepared by depositing 6 µm of SiO<sub>2</sub> on a silicon substrate. Then this sample is polished from the side, and 0.115 µm of Si<sub>3</sub>N<sub>4</sub> is deposited on the polished side. The refractive index and thickness of the Si<sub>3</sub>N<sub>4</sub> layer were determined with a conventional ellipsometer by measuring a reference silicon substrate placed beside the sample during the deposition. The cross section of this sample is illustrated in Fig. 15. The ellipsometric image and its cross section obtained by the imaging microellipsometer are shown in Fig. 16. The polarization contrast obtained in this case comes from the index of refraction variations beneath the Si<sub>3</sub>N<sub>4</sub> layer. From the result, it is obvious that the response of different refractive index regions is distinct. However, the index of refraction cannot be accurately determined from the ratio signal. This is attributed to the fact that the measured properties of the Si<sub>3</sub>N<sub>4</sub> on the reference silicon wafer do not represent the true properties of the Si<sub>3</sub>N<sub>4</sub> on the sample. To estimate the accuracy of the refraction index measurement, the refraction index of the Si<sub>3</sub>N<sub>4</sub> on the reference sample is measured by using the imaging ellipsometer. Given the thickness, the index of refraction was determined to be 1.992, compared to 1.932 measured by a conventional ellipsometer. The imperfection of the polish and the low reflectivity in this experiment are the major sources of noise. From Fig. 16b, the noise level of the ratio signal is determined to be ±0.005. Based on this noise level, the sensitivity to the index of refraction is estimated to be ±0.03. With a brighter source and better polish, a sensitivity much better than ±0.01 is expected.

**Fig. 14** Imaging ellipsometric measurement of the PMMA grating.  
**(a)** Ellipsometric ratio image.  
**(b)** Line scan of **(a)** after being converted into thickness

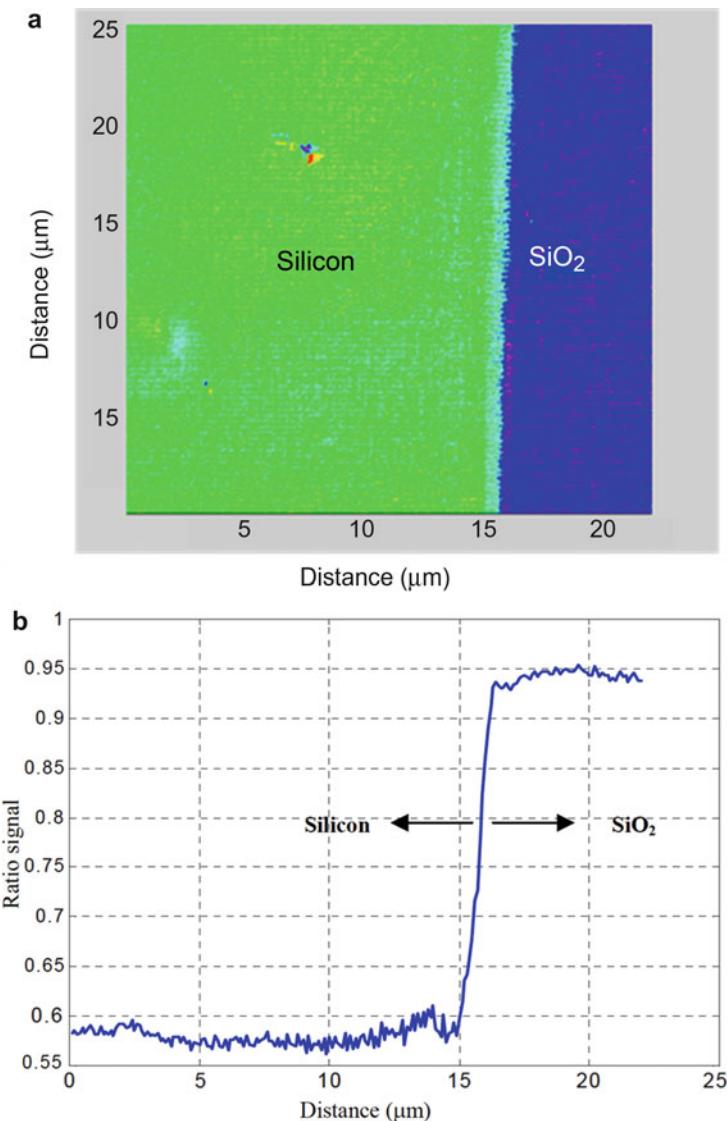


**Fig. 15** Cross section of the sample used in the refractive index measurement



### Measurement of Surface Features Beyond the Diffraction Limit

Optical characterization techniques have long been indispensable tools for advanced micro-fabrication. However, the ever-decreasing minimum feature size has pushed many optical techniques to the limit of their capabilities. Owing to the



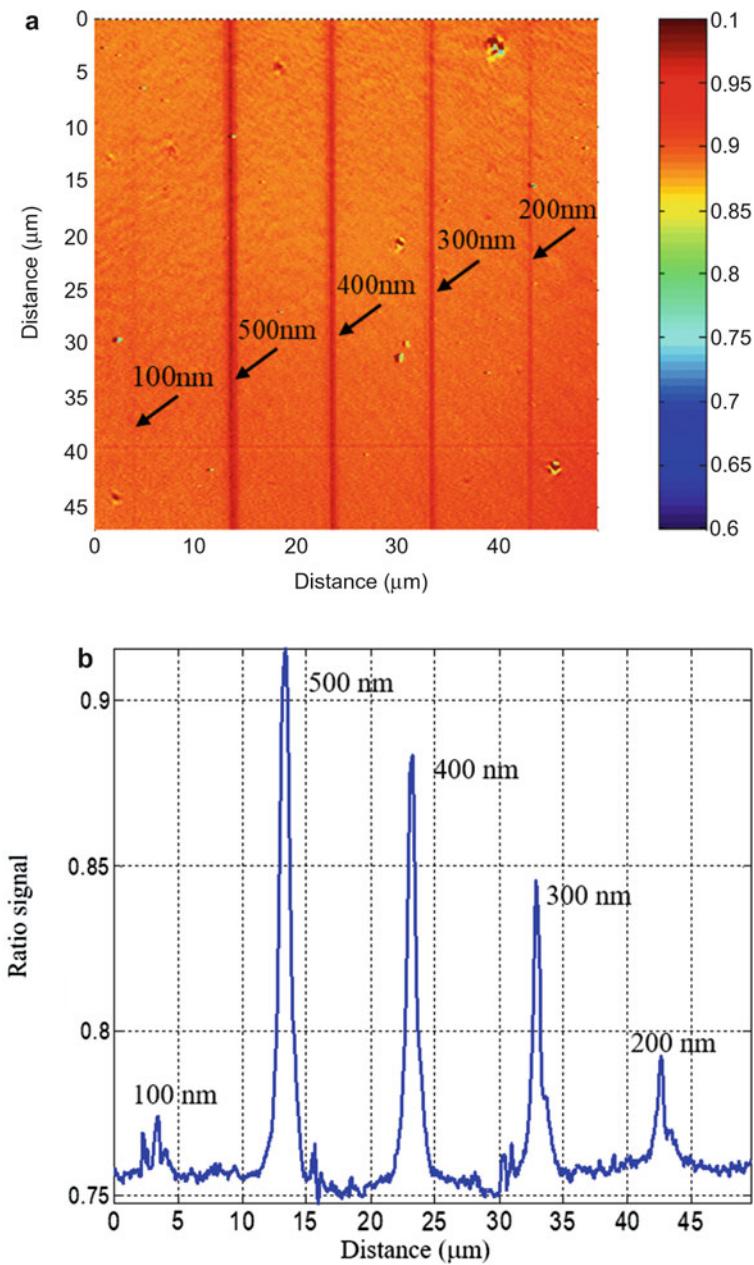
**Fig. 16** Experimental results of index measurement. (a) Ellipsometric ratio image. (b) Line scan of (a)

diffraction phenomena, it is difficult for optical techniques to resolve features with sub-wavelength lateral dimensions. Near-field techniques, such as near-field scanning optical microscopy (NSOM) and photon tunneling microscopy (PTM), have been researched to improve the resolution of optical systems. However, it is also possible to characterize sub-wavelength features using far-field techniques without actually resolving them. Provided with certain a priori information, which is often

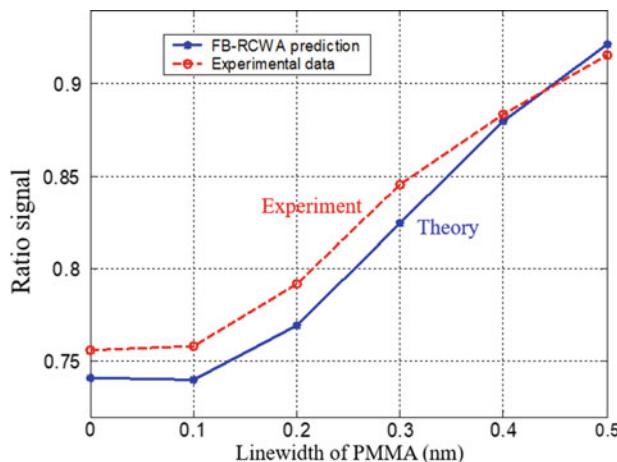
available in modern micro-fabrication, characterization of sub-wavelength structures can be performed through parametric model fitting. For example, scatterometry has been used to characterize periodic sub-wavelength structures by investigating the far-field scattering signature (Minhas et al. 1998). It is well known that sub-wavelength structures can alter the polarization of incident light. In optical metrology applications, this effect can serve as a sensitive measurement of the sub-wavelength structures, which has been demonstrated in periodic structure characterization (Minhas et al. 1998) and isolated trench measurement (Wang et al. 2001). Using the imaging microellipsometer, the application of this polarization effect is extended to characterize isolated sub-wavelength structures in a parallel fashion (Zhan and Leger 2001, 2002e).

The imaging ellipsometric signals of a sample containing submicron lines are measured to demonstrate the polarization effects from sub-wavelength structures and their application to characterize unresolved features. The sample is prepared with e-beam lithography. A 105 nm SiO<sub>2</sub> layer was deposited on a silicon wafer followed by a 50 nm layer of PMMA. The PMMA was removed by e-beam lithography to form lines with 100, 200, 300, 400, and 500 nm widths. Figure 17a presents the experimental measurement of the ratio image along with a line scan of the ellipsometric image shown in Fig. 17b. The experimental result clearly shows the polarization effect from sub-wavelength features. Different linewidths give different polarization ratio signals. This polarization ratio signal can be regarded as a sensitive signature of linewidth. The measurement is repeated on similar test lines, and the previously obtained ratio signals are used as calibrated signals to calculate the linewidths. From the results, the accuracy of the measurement has been determined to be about 10 nm for lines that are broader than 100 nm. The signal for 100 nm line is more susceptible to noise, but the signal has been clearly observed. To verify the signal indeed comes from the polarization effect, the exactly same procedures were performed without polarizers in the setup. The ratio image obtained in this manner was essentially unity, and the signatures of the submicron lines disappeared.

A focused beam rigorous coupled wave analysis (FB-RCWA) (Otaki et al. 2000) has been developed to study the polarization effects from nonperiodic sub-wavelength structures. The RCWA method is inherently designed for periodic structures (Moharam et al. 1995a, b). However, if the period of the structure is large enough and a small spot illumination is synthesized with multiple plane waves, the optical response from a nonperiodic structure can be obtained. FB-RCWA simulations of these sub-wavelength structures were performed, and the corresponding ratio signals were calculated and shown in Fig. 18 along with the experimental data. Despite a small discrepancy between theoretical and experimental data, the tendency of experimental signals is consistent with the simulation. The FB-RCWA simulation also predicts the dip at the center of 100 nm line shown in Fig. 17b (Zhan and Leger 2002f). Considering the complex physics that governs this problem, the match between the experimental curve and theoretical curve calculated by the simple optical model is remarkably good. The theoretical calculations are sensitive to many other parameters, such as residual birefringence of the objective lens, the retardation from the wave plate, the index of refraction of each layer, the thickness of the film, and so



**Fig. 17** Experimental results of submicron PMMA lines measurement. **(a)** Ratio signal image. **(b)** Line scan of **(a)**

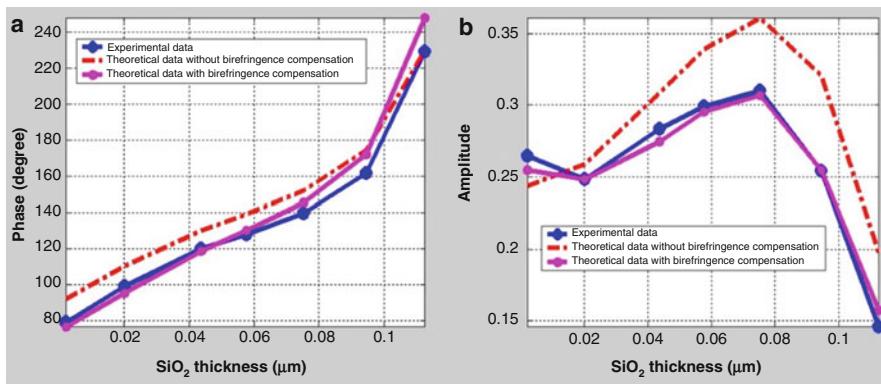


**Fig. 18** Comparison between theoretical and experimental imaging microellipsometric signals for unresolved lines

on. In practice, it is unnecessary to obtain an exact match between the two curves in Fig. 18. The accuracy of an actual measurement tool would be improved substantially by compensating these effects through a calibration procedure.

### Measurement of Axial Birefringence

Axial birefringence is birefringence distribution that obeys rotational symmetry. This birefringence often occurs in objective lens due to optical coating and stress. This birefringence causes polarization aberrations that degrade the performance of the optical devices (Stallinga 2001). Accurate characterization of this axial birefringence is very important for the applications of these optical devices to photolithography and optical data storage. The rotationally symmetric microellipsometer can be employed to accurately characterize the axial birefringence in optical systems and devices. In this experiment, a series of  $\text{SiO}_2$  thin films on silicon with different thickness were used. The  $\text{SiO}_2$  was deposited onto silicon substrates by the use of PECVD (Plasma-Enhanced Chemical Vapor Deposition), and the thickness and refractive index were measured with a spectroscopic ellipsometer. The measured thickness and index were then used in the optics model to calculate the amplitude ( $\cos 2\epsilon$ ) and phase ( $2\theta_0$ ) of the signal. The calculated values were then compared with the experimental results. An appreciable discrepancy between the experimental and the theoretical data was found with the assumption that the objective lens was perfect. A small amount of residual axial birefringence was discovered in the used objective. After including this residual birefringence ( $2^\circ$  retardation) from the objective lens in the optical model, very good agreement between the theoretical and experimental

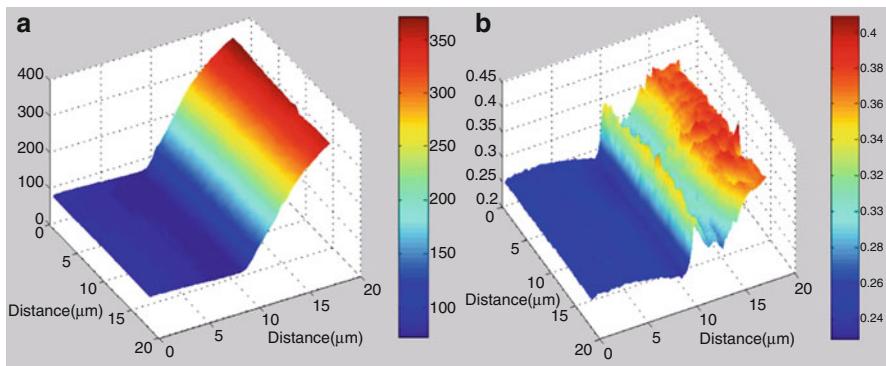


**Fig. 19** Experimental and theoretical data for a series of  $\text{SiO}_2$  thin films on silicon substrate. (a) Phase. (b) Amplitude

curves was obtained. The experimental and theoretical data are shown in Fig. 19. The measured amount of axial birefringence is comparable to those reported in literature (Shribak et al. 2002).

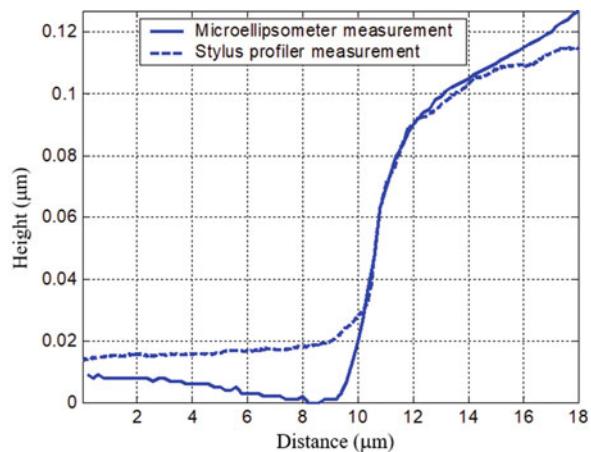
## Measurement of Micro-optical Components

Recently, micro-optical components became increasingly important due to their wide applications in communications and in information technology. Accurate characterization of these devices plays a critical role in successful mass production of such devices. The rotationally symmetric microellipsometer can be used to characterize micro-optical components. A photoresist microprism is fabricated with gray-scale mask technique. Ellipsometric images of this microprism are obtained through scanning in two dimensions (as shown in Fig. 20). The image size is  $18 \times 18 \mu\text{m}$  with  $0.2 \mu\text{m}$  scanning step size. A resolution of  $0.5 \mu\text{m}$  can be estimated from the illumination wavelength and numerical aperture. The integration time of the lock-in was set to be 100 ms. Due to the vibration caused by the motor, the amplitude data was relatively noisy compared to the phase data. The refractive index of photoresist is measured with a standard ellipsometer, and then this refraction index and the phase data are utilized to calculate the thickness of the photoresist microstructure. A line scan is shown in Fig. 21. Despite some deviations, this result was in good agreement with a profiler measurement. The deviations could be caused by the fact that the two instruments don't measure the exactly same location of the sample. In addition, it is noteworthy that the profiler measurements show a slight increase in thickness for the leftmost part of the line scan, while the microellipsometer data shows a slight decrease of thickness for the same region. This is due to the large stylus size ( $\sim 5 \mu\text{m}$  radius) and the leveling error of the profiler. The comparison clearly demonstrates the advantages of the microellipsometer over a stylus profiler. Besides the noncontact and nondestructive characteristics, the microellipsometer measurement gives the absolute



**Fig. 20** Experimental results of a photoresist microprism measured by the rotationally symmetric microellipsometer. (a) Phase image. (b) Amplitude image

**Fig. 21** Comparison of surface profile measured by the microellipsometer and a stylus profiler



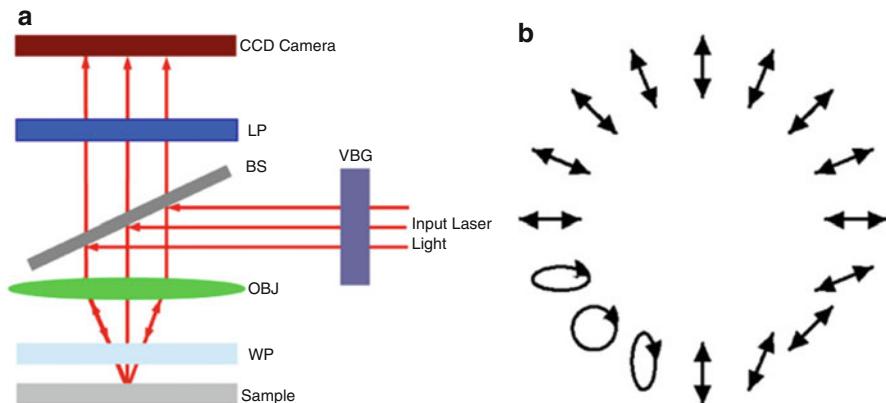
thickness of the sample, instead of the topology, with higher spatial resolution. It is necessary to point out that, although only the thickness was measured in this experiment, simultaneous determination of thickness and refractive index of sample is possible with less noisy amplitude data, which is another advantage of the microellipsometer.

## Rapid Mueller Matrix Polarimetry

Another recent development in optical polarization-based metrology is a rapid Mueller matrix polarimetry that takes advantage of the recent rapid progress made in vectorial optical fields. Just like ellipsometry, polarimetry refers to an optical measurement technique that utilizes the detection of polarization changes upon

light-matter interactions for the applications of optical characterization, imaging and sensing. The main difference relies in the fact that ellipsometry mostly deals with totally polarized and coherent light while partially polarized, while polarimetry can handle both totally coherent and partially coherent optical fields. In order to handle partially polarized and partially coherent light fields, Stokes parameters are used in polarimetry to describe the light polarization. Consequently, the polarization responses of samples and testing materials are represented by a  $4 \times 4$  Mueller matrix, and the characterization task is turned into the determination of the 16 elements of this matrix. Determination of the elements of the Mueller matrix in a Mueller matrix polarimetry (MMP) is done by analyzing the reflected or transmitted optical polarization states with several known input polarization states created by a polarization state generator (PSG). Traditionally, these known polarization states are created in a sequential manner with the use of rotating or modulating optical devices, which ultimately limit the speed of a MMP and its applications in real-time characterization of dynamic events.

The developments in vectorial optical fields offer a unique opportunity to improve the speed of the MMP through parallelizing the PSG. One of such examples is illustrated in Fig. 22 (Tripathi and Toussaint 2009). An input vectorial optical field with spatially designed known polarization states is used as the illumination at the input pupil of an objective lens. The objective lens focuses the input vectorial optical field onto the sample through a  $\lambda/4$ -plate, and the reflected signal passes through the same  $\lambda/4$ -plate, objective lens, and is projected onto a CCD camera after passing through a linear polarizer. The resulting intensity patterns are used for the determination of the Mueller matrix of the sample under test. Through the use of a vectorial optical field as the input, the PSG function is parallelized, and a rapid MMP measurement is realized with a single shot. The performance of this rapid MMP is numerically investigated with the use of anisotropic samples such as anisotropic thin film and stratified metal-dielectric metamaterials.



**Fig. 22** (a) Schematic diagram of a rapid MMP and (b) the polarization state distribution of the incident vectorial optical field

## Summary

In this chapter, two spatially resolved microellipsometry techniques and their applications in characterization of patterned surface structures have been reviewed. Meanwhile, based on polarization diversities within the cross section of vectorial optical fields, rapid Mueller matrix polarimetry technique has also been described with single snapshot measurement. These techniques have wide applications in the materials research, in the semiconductor and microelectronics industry, in the optics and photonics industry, and in many other disciplines. Based on the microellipsometry techniques described above, current research is ongoing to design spectroscopic microellipsometers and generalized microellipsometers. These techniques will significantly expand the capabilities of ellipsometry and enables accurate characterization and diagnostics of more complicated materials. The resolution of these microellipsometers can be further increased by incorporating near-field optical techniques (such as near-field optical microscopy (NSOM) and solid immersion lens (SIL) (Zhan and Leger 2003)). Cylindrical vector beams may also be incorporated in the rotationally symmetric microellipsometer and MMP to further increase their spatial resolution (Zhan and Leger 2002c; Quabis et al. 2000). Furthermore, through integrating other optical and nonoptical techniques (such as reflectometry, optothermal techniques, acoustics, etc.) with the microellipsometry, one can obtain more information about the sample and more data to analyze, allowing more accurate diagnostics and characterization of materials and structures with unprecedented spatial resolution.

---

## References

- Albersdörfer A, Elender G, Mathe G, Neumaier K R, Paduschek P, Sackmann E (1998) High resolution imaging microellipsometry of soft surfaces at 3  $\mu\text{m}$  and 5 Å normal resolution. *Appl Phys Lett* 72:2930–2932
- Azzam RMA, Bashara NM (eds) (1977) Ellipsometry and polarized light. North Holland Publishing Company, New York
- Erman M, Theeten JB (1986) Spatially resolved ellipsometry. *J Appl Phys* 60:859–873
- Gold N, Willenborg DL, Opsal J, Rosencwaig A (1989) High resolution ellipsometric apparatus. US Patent #5042951
- Herman IP (ed) (1996) Optical diagnostics for thin film processing. Academic, San Diego
- Jin G, Jansson R, Arwin H (1996) Imaging ellipsometry revisited: development for visualization of thin transparent layers on silicon substrates. *Rev Sci Instrum* 67:2930–2935
- Leng JM, Chen J, Fanton J, Senko M, Ritz K, Opsal J (1998) Characterization of titanium nitride (TiN) films on various substrates using spectrophotometry, beam profile reflectometry, beam profile ellipsometry and spectroscopic beam profile ellipsometry. *Thin Solid Films* 313:308–313
- Liu A, Wayner PC, Plawsky JL (1994) Image scanning ellipsometry for measuring nonuniform film thickness profiles. *Appl Opt* 33:1223–1229
- Minhas BK, Coulombe SA, Naqvi SSH, McNeil JR (1998) Ellipsometric scatterometry for the metrology of sub-0.1- $\mu\text{m}$ -linewidth structures. *Appl Opt* 37:5112–5115
- Moharam MG, Grann EB, Pommet DA, Gaylord TK (1995a) Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings. *J Opt Soc Am A* 12:1068–1076

- Moharam MG, Grann EB, Pommet DA, Gaylord TK (1995b) Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach. *J Opt Soc Am A* 12:1077–1086
- Otaki K, Osawa H, Ooki H, Saito J (2000) Polarization effect on signal from optical ROM using solid immersion lens. *Jpn J Appl Phys* 39:698–706
- Quabis S, Dorn R, Eberler M, Glöckl O, Leuchs G (2000) Focusing light into a tighter spot. *Opt Commun* 179:1–7
- Richards B, Wolf E (1959) Electromagnetic diffraction in optical systems II. Structure of the image field in an aplanatic system. *Proc R Soc Ser A* 253:358–379
- See CW, Somekh MG, Holmes RD (1996) Scanning optical microellipsometer for pure surface profiling. *Appl Opt* 35:6663–6668
- Shribak M, Inoué S, Oldenbourg R (2002) Polarization aberrations caused by differential transmission and phase shift inn high-numerical-aperture lenses: theory, measurement, and rectification. *Opt Eng* 41:943–954
- Stallinga S (2001) Axial birefringence in high-numerical-aperture optical systems and the light distribution close to focus. *J Opt Soc Am A* 18:2846–2859
- Tompkins HG (ed) (1993) A users' guide to ellipsometry. Academic, New York
- Tompkins HG, McGahan WA (eds) (1999) Spectroscopic ellipsometry and reflectometry: a user's guide. Wiley, New York
- Tripathi S, Toussaint KC (2009) Rapid Mueller matrix polarimetry based on parallelized polarization state generation and detection. *Opt Express* 17:21396–21407
- Tschimwang A, Zhan Q (2010) High-spatial-resolution nulling microellipsometer using rotational polarization symmetry. *Appl Opt* 49:1574–1580
- Wang X, Mason J, Latta M, Strand T, Marx D, Psaltis D (2001) Measuring and modeling optical diffraction from subwavelength features. *J Opt Soc Am A* 18:565–572
- Wolf E (1959) Electromagnetic diffraction in optical systems I. An integral representation of the image field. *Proc R Soc Ser A* 253:349–357
- Zhan Q, Leger JR (2001) Imaging ellipsometry for high-spatial-resolution metrology. *Invited paper.* SPIE Proc 4435:65–76
- Zhan Q, Leger JR (2002a) A high resolution imaging ellipsometer. *Appl Opt* 41:4443–4450
- Zhan Q, Leger JR (2002b) Interferometric measurement of Berry's phase in space-variant polarization manipulations. *Opt Commun* 213:241–245
- Zhan Q, Leger JR (2002c) Microellipsometer with radial symmetry. *Appl Opt* 41:4630–4637
- Zhan Q, Leger JR (2002d) Focus shaping using cylindrical vector beams. *Opt Express* 10:324–331
- Zhan Q, Leger JR (2002e) Measurement of surface features beyond the diffraction limit using an imaging ellipsometer. *Opt Lett* 47:821–823
- Zhan Q, Leger JR (2002f) Measuring unresolved surface features using imaging ellipsometric polarization signatures. In: DOMO technical digest, Diffractive-Optics Micro-optics (DOMO) topical meeting, Tucson, pp 101–103
- Zhan Q, Leger JR (2003) Near-field nano-ellipsometer for ultrathin film characterization. *J Microsc* 210:214–219



# Raman Scattering

17

Rudolph M. Erasmus and J. Darrell Comins

## Contents

Introduction .....	542
Basics of Raman Scattering .....	543
The Raman Effect .....	543
Instrumentation .....	558
Light Source .....	559
Sample Considerations .....	560
Collection Optics .....	560
Post-Sample Filtering .....	562
Polarization Control .....	562
Wavelength Selector .....	562
Detector .....	563
Advantages and Disadvantages of Raman Spectroscopy .....	563
Typical Applications .....	564
Determination of Stress .....	565
Applications Using Phase Identification .....	584
Comments on Other NDE Applications and Technique Developments .....	590
Summary .....	591
References .....	592

---

R. M. Erasmus

School of Physics, Materials Physics Research Institute and DST-NRF Centre of Excellence in Strong Materials, University of the Witwatersrand, Johannesburg, South Africa

Microscopy and Microanalysis Unit, University of the Witwatersrand, Johannesburg, South Africa  
e-mail: [rudolph.erasmus@wits.ac.za](mailto:rudolph.erasmus@wits.ac.za)

J. D. Comins (✉)

School of Physics, Materials Physics Research Institute and DST-NRF Centre of Excellence in Strong Materials, University of the Witwatersrand, Johannesburg, South Africa  
e-mail: [darrell.comins@wits.ac.za](mailto:darrell.comins@wits.ac.za)

**Abstract**

Raman scattering is an inelastic light-scattering technique that finds wide application in physics, chemistry, geology, engineering, and life sciences. It is a nondestructive evaluation method that gives information on vibrational modes in sample materials and can thus be used for characterization of structure and composition of materials. This includes phase identification (study of polytypes and phase transitions), characterization of residual stress and strain, studies of nanomaterials, radiation damage, catalysis processes, corrosion mechanisms, oxide formation, metabolic process in biological tissues, and many more. Information is obtained at the micron scale in materials, and Raman mapping can be used to determine composition and stress/strain in materials at similar spatial scales. Advances in instrumentation over the past decade or so have made the technique more widely accessible, and a brief overview of dispersive Raman instrumentation is given. The description of the basics of the method focuses on solid materials, and applications focus on phase identification and NDE of stress and strain. The examples covered in more detail include stress determination in diamond and polycrystalline diamond (PCD) tools and *in situ* characterization of corrosion processes on iron surfaces. The chapter concludes with a short overview of other NDE applications, with citations of relevant literature, and method developments such as tip-enhanced Raman spectroscopy (TERS).

---

**Introduction**

One of the avenues of investigation open to an experimentalist to study the detailed structure of matter is the study of the interaction of light (photons) and matter (gases, liquids, and solids). Broadly speaking, this interaction can result in the absorption, emission, or scattering of the incident electromagnetic radiation. These phenomena are intimately related to transitions between the energy levels in the matter, be it in the form of atoms, molecules, or crystals. Transitions between rotational energy levels are associated with electromagnetic radiation in the microwave region of the spectrum, vibrational energy levels with infrared wavelengths, and transitions between electronic levels with optical and ultraviolet wavelengths. A transition from a lower level to a higher level with transfer of energy from the radiation field to the host matter is called absorption. A transition from a higher to a lower level is called emission if energy is transferred to the radiation field or non-radiative decay if no radiation is emitted, e.g., decay via phonons in a crystal lattice. Scattering (or redirection) of light due to interaction with matter may or may not occur with transfer of energy, i.e., the scattered radiation has, respectively, a slightly different wavelength or the same wavelength. In this chapter, we are chiefly concerned with the **Raman effect**, which is an inelastic scattering process, i.e., it takes place with some form of transfer of energy. Brief mention will also be made of **photoluminescence**, which is an emission process that can take place simultaneously with Raman scattering.

Raman scattering is of interest as a nondestructive evaluation (NDE) technique for a variety of reasons. It provides relatively fast phase identification at the micron scale and can be used to determine the presence of stress and strain in materials at similar spatial scales.

Provided certain precautions are taken, Raman scattering is almost always non-destructive in nature. In order to limit the scope of this discussion, the description of the basics of the method focuses on solid materials and applications focus on phase identification and NDE of stress and strain. The chapter concludes with a short overview of other NDE applications and method developments such as tip-enhanced Raman spectroscopy (TERS).

In terms of general references for the technique and its applications, including discussions on instrumentation, the reader is referred to *Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line*, edited by Ian R. Lewis and Howell G.M. Edwards (Marcel-Dekker, 2001), and *Handbook of Vibrational Spectroscopy* (5 volumes), edited by John M. Chalmers and Peter R. Griffiths (J Wiley, 2002). The latter also includes discussions on infrared absorption spectroscopy.

The *Journal of Raman Spectroscopy* publishes a very useful and comprehensive annual review that discusses recent advances in linear and non-linear Raman spectroscopy. The most recent such review, Part XI, was authored by Nafie (2017). It is shown that the number of papers published involving Raman spectroscopy from 2007 to 2016 has just more than doubled in this time period. Earlier parts of these reviews (I to X from 2007 onward) can be found in the same journal.

---

## Basics of Raman Scattering

### The Raman Effect

In the writing of this section, various texts were consulted. Principal use was made of Brüesch (1986) Reprinted/adapted by permission from Springer Customer Service Centre GmbH: Springer, Phonons: Theory and Experiments II by Peter Brüesch, COPYRIGHT 1986. The reader is also referred to Griffith (1975), Thorne (1988), Klein (1990), Long (2002), and Keresztfury (2002) for general overviews and the book by Marcuse (1980) for a comprehensive quantum electronic description of the Raman effect.

### Introduction

The possible existence of the Raman effect was predicted in 1923 by Adolf Smekal, and in 1928, Sir Chandrasekhara V. Raman, using sunlight filtered through a prism and toluene as the scattering medium, demonstrated the phenomenon to which his name was subsequently given (Raman 1928; Raman and Krishnan 1928). It is sometimes referred to as the Smekal-Raman effect by continental writers. Independently, and only a month later, Landsberg and Mandelstam (1928a, b) observed the

phenomenon using mercury radiation on quartz and so were the first to observe the Raman spectrum of a mineral.

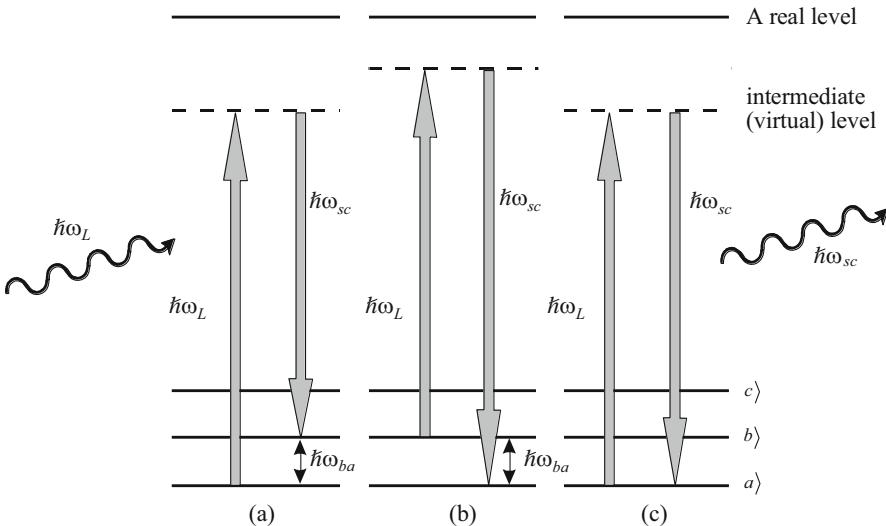
In the following, the discussion focuses primarily on solids. When matter is irradiated by non-resonance monochromatic radiation of frequency  $\nu$ , a certain fraction of the radiation is scattered. Light scattering arises from three main factors, named after their discoverers, namely, Tyndall, Rayleigh, and Raman. Tyndall scattering is caused by the presence of very small but macromolecular particles in the substance under illumination and does not involve a change in the frequency of the scattered light. Both Rayleigh and Raman scattering derive from the interaction of the radiation with the crystal lattice of the sample. The Rayleigh effect, where the light is scattered elastically (the direction of the incident photons is changed, but not their wavelength), accounts for the majority of the scattered light and also does not involve a change in the frequency of the scattered light. The intensity of the Raman light is five to six orders of magnitude less than that of the Rayleigh scattered light and does involve a change in the frequency of the scattered light due to the light being scattered inelastically. The Raman spectrum of a specific sample shows a number of discrete scattered frequencies both higher and lower than that of the exciting radiation ( $\nu$ ). Such new frequencies in the scattered light are the so-called Raman bands and each has an associated Raman shift ( $\Delta\nu$ ) relative to the exciting radiation frequency  $\nu$ . Those bands of lower frequency than the exciting radiation are called Stokes lines and are usually stronger than those bands of higher frequency than the exciting radiation, the anti-Stokes lines. In first-order scattering, only one phonon is involved; this corresponds to the terms linear in the normal coordinates  $Q_s$  (Eq. 17). In second-order scattering, two phonons are involved, corresponding to the terms proportional to  $Q_s' Q_{s''}$  in Eq. 17 or to anharmonic coupling of a phonon which is active in first-order Raman scattering with two other phonons.

There are many different ways in which the effect can be described. A rigorous semiclassical treatment of the Raman effect requires third-order perturbation theory while its quantum-mechanical description as a two-photon process requires second-order perturbation theory (Marcuse 1980). In solids, it is necessary to explicitly describe the vibrational waves which are excited by the Raman process, so that third-order perturbation theory is again necessary.

In the following paragraphs, both a simple classical and quantum-mechanical treatment of the Raman process is given that suffices to describe the observed features of a Raman spectrum and facilitates the understanding of the Raman process on a microscopic level.

### A Simple Quantum Mechanical Approach

This subsection gives an introductory qualitative discussion in terms of discrete energy levels and of photons. The quantum-mechanical model is the two-photon process shown schematically in Fig. 1. Let  $(\omega_L, \mathbf{k}_L)$  be the incident photon of the laser with frequency  $\omega_L$  and wave vector  $\mathbf{k}_L$ ,  $(\omega_{sc}, \mathbf{k}_{sc})$  the scattered photon, and  $(\omega_j, \mathbf{q})$  the optical phonon  $s = (\mathbf{q}, j)$  involved in the scattering process. Energy and momentum are conserved between initial and final states of the system. For Rayleigh scattering, we have  $\omega_L = \omega_{sc}$  and  $\mathbf{k}_L = \mathbf{k}_{sc}$ . For Raman scattering, the



**Fig. 1** Schematic energy level diagram illustrating the (a) Stokes and (b) anti-Stokes transitions that can occur during the Raman process. (c) is the case for Rayleigh scattering

conservation of energy and momentum are  $\omega_L = \omega_{sc} \pm \omega_j(\mathbf{q})$  and  $\mathbf{k}_L = \mathbf{k}_{sc} \pm \mathbf{q}$ , respectively. Since  $\omega_L \gg \omega_s = \omega_j(\mathbf{q})$ , it follows from energy conservation that  $\omega_L \cong \omega_{sc}$ . Experiments are usually carried out at frequencies where there is essentially no dispersion of the refractive index  $n$ , i.e., the transparent region of the crystal. Since  $\mathbf{k}_L$  and  $\mathbf{k}_{sc}$  are the wave vectors within the crystal, we have  $k_L = 2\pi/\lambda_L$ ,  $k_{sc} = 2\pi/\lambda_{sc}$ , where  $\lambda_L = \lambda_v/n(\omega_L)$  and  $\lambda_{sc} = \lambda_v/n(\omega_{sc})$ , with  $\lambda_v$  the wavelength in vacuum. From  $c = \nu\lambda_v$ , we obtain  $k_L = n(\omega_L)\omega_L/c$ ,  $k_{sc} = n(\omega_{sc})\omega_{sc}/c$ , and since  $\omega_L \cong \omega_{sc}$ , it follows that  $k_L \cong k_{sc}$ . In addition,  $\lambda_L$  and  $\lambda_{sc}$  are much larger than the lattice parameter  $a$ , and hence,  $k_L$  and  $k_{sc}$  are much smaller than  $\pi/a$ , the magnitude of the wave vector at the zone boundary. Thus from momentum conservation, it follows that  $q \ll \pi/a$ , i.e., in first-order Raman scattering, only  $\mathbf{q} \cong 0$  optical modes can be excited.

With reference to Fig. 1, suppose that we denote some vibrational state of a crystal as  $|a\rangle$ . This need not necessarily be an excited state. An incident photon of energy  $\hbar\omega_L$  is absorbed, raising the system to some intermediate or virtual state. Insofar as this is energetically distant from a real level (large  $\delta E$ ), the uncertainty principle,  $\delta E\delta t \sim \hbar$ , allows the virtual state a vanishingly small lifetime (of the order of  $10^{-14}$  seconds). The system immediately makes a Stokes transition, emitting a (scattered) photon of energy  $\hbar\omega_{sc} < \hbar\omega_L$ . In conserving energy, the difference  $\hbar\omega_L - \hbar\omega_{sc} = \hbar\omega_{ba}$  goes into exciting the crystal to a higher vibrational energy level  $|b\rangle$ . It is possible that electronic or rotational excitation results as well. Alternatively, if the initial state is an excited one, the crystal, after absorbing and emitting a photon, may drop back to an even lower state  $|a\rangle$ , thereby making an anti-Stokes transition. In this instance,  $\hbar\omega_{sc} > \hbar\omega_L$ , which means that some vibrational energy of the crystal ( $\hbar\omega_{sc} - \hbar\omega_L = \hbar\omega_{ba}$ ) has been converted to radiant energy. In either case, the resulting differences between

$\omega_{sc}$  and  $\omega_L$  correspond to specific energy level differences for the substance under study and as such yield insights into crystal structure.

It is important to note that the “absorption” of the photon  $\hbar\omega_L$  and the emission of the photon  $\hbar\omega_{sc}$  are truly simultaneous events and cannot be separated in time from one another. The apparent time sequence of events in the above description is merely part of the process of explanation.

This description of the Raman effect accounts for the observed difference in the intensities of the Stokes and anti-Stokes bands, since the population of the ground vibrational level (Fig. 1a) is much greater, for optical phonons and not too high temperatures, than the population of excited vibrational levels (Fig. 1b). Thus the chance the incident photon finds the system in an excited vibrational level is much smaller than for the ground state. Since the ratio of the two populations, according to Bose-Einstein statistics, is proportional to  $\exp[-\hbar\omega_s/k_B T]$ , the ratio of the intensities of a Stokes line to a corresponding anti-Stokes line is expected to be proportional to  $\left(\frac{\omega_L - \omega_s}{\omega_L + \omega_s}\right)^4 \exp\left(\frac{\hbar\omega_s}{k_B T}\right)$  and this ratio is considerably larger than unity.

### A Simple Classical Approach

For this discussion, we consider a single molecule, but it is equally applicable to an inter-atomic bond in a solid. When electromagnetic radiation impinges on such a sample, the electric field  $\mathbf{E}$  induces an atomic dipole, the nuclei of the solid in the vicinity of the  $\mathbf{E}$  field being displaced toward the negative end of the dipole and the electron cloud to the positive part. Let  $\mathbf{E} = E_0 \cos \omega_L t$  be the electric vector of the incident light. Since  $\omega_L$  is much larger than the vibrational frequencies of the atoms, only the electrons but not the atoms can respond to the light field  $\mathbf{E}$ . For a dipole moment  $\mathbf{M}$  induced by  $\mathbf{E}$ , we write

$$\mathbf{M} = \alpha \mathbf{E} + \frac{1}{2} \beta \mathbf{E}^2 + \dots, \quad (1)$$

where  $\alpha$  is the electronic polarizability of the sample matter and can be thought of as being related to the ease with which the electron cloud may be deformed. Both  $\mathbf{M}$  and  $\mathbf{E}$  are vectors,  $\alpha$  is a second-order tensor, and  $\beta$  is a third-order tensor known as the hyperpolarizability, which gives rise to the hyper-Raman effect. This effect will not be discussed here.

In general, the direction of  $\mathbf{M}$  does not coincide with the direction of  $\mathbf{E}$ , i.e.,  $\alpha$  is a second-order tensor with components  $\alpha_{ij}$ . For the present discussion, it is assumed that  $\mathbf{M}$  is parallel to  $\mathbf{E}$  (this is the case for isotropic systems or if  $\mathbf{E}$  is parallel to the direction of one of the axes of high symmetry of the system). The electronic polarizability  $\alpha$  depends on the electric charge distribution of the system. If the atomic configuration changes during a vibration, the charge distribution and hence  $\alpha$  will also change. Expanding  $\alpha$  in a Taylor series gives

$$\alpha = \alpha_0 + \left( \frac{\partial \alpha}{\partial Q} \right)_0 Q + \frac{1}{2} \left( \frac{\partial^2 \alpha}{\partial Q^2} \right)_0 Q^2 + \dots, \quad (2)$$

with  $Q$  a normal coordinate and  $\alpha_0$  the polarizability of the molecule when it is not vibrating. The first-order Raman effect is determined by the term linear in  $Q$ , the second-order Raman effect by the term quadratic in  $Q$ , etc. In the following, only first-order Raman scattering is considered. If a particular bond in the molecule vibrates with the frequency  $\omega_s$ , then  $Q = Q_0 \cos \omega_s t$  and

$$\alpha(t) = \alpha_0 + \left( \frac{\partial \alpha}{\partial Q} \right)_0 Q_0 \cos \omega_s t. \quad (3)$$

Substituting Eq. 3 into Eq. 1 gives

$$\mathbf{M}(t) = \alpha_0 \mathbf{E}_0 \cos \omega_L t + \left( \frac{\partial \alpha}{\partial Q} \right)_0 Q_0 \mathbf{E}_0 \cos \omega_L t \cos \omega_s t. \quad (4)$$

On expansion, this gives

$$\mathbf{M}(t) = a \cos \omega_L t + b [\cos(\omega_L - \omega_s)t + \cos(\omega_L + \omega_s)t], \quad (5)$$

where

$$a = \alpha_0 \mathbf{E}_0; \quad b = \frac{1}{2} \left( \frac{\partial \alpha}{\partial Q} \right)_0 Q_0 \mathbf{E}_0 \quad (6)$$

Equation 5 shows that the induced dipole moment  $\mathbf{M}$  vibrates not only with the frequency  $\omega_L$  of the incident light but also with the frequencies  $\omega_L \pm \omega_s$ . Thus we have three frequency components. The first term of Eq. 5 is the classical expression of Rayleigh scattering since an oscillating dipole will itself emit radiation of frequency  $\omega_L$ , and the intensity of the radiation is proportional to the mean square of  $\ddot{\mathbf{M}}(t)$ . The second and third terms give the Stokes and anti-Stokes Raman components, respectively.

An important consequence of Eq. 5 is that it also gives the basic selection rules for the activity of a Raman vibration. If the rate of change of polarizability with the bond vibration  $\omega_s$  is zero, then that vibration will be Raman inactive and there will be Rayleigh scattering only. Thus for a vibration to be Raman active, a change of polarizability must be involved.

From classical radiation theory of an oscillating dipole, the intensity of the emitted radiation by the dipole moment  $\mathbf{M}(t)$  into the solid angle  $d\Omega = \sin \vartheta d\vartheta d\varphi$  is given by (Crawford 1968; Brüesch 1986)

$$dI(t) = \frac{d\Omega}{4\pi c^3} \sin^2 \vartheta |\ddot{\mathbf{M}}(t)|^2 \quad (7a)$$

and per unit solid angle  $d\Omega$  by

$$I(t) = \frac{1}{4\pi c^3} \sin^2 \vartheta |\ddot{\mathbf{M}}(t)|^2 = A |\ddot{\mathbf{M}}(t)|^2. \quad (7b)$$

In MKS units, the intensity or time-averaged power per unit solid angle radiated by an oscillating electric dipole induced by the electric field of the incident radiation of frequency  $\omega_L$  along a direction making an angle  $\vartheta$  with the axis of the dipole is given by

$$I = \frac{\omega_s^4 Q_0^2}{32\pi^2 \epsilon_0 c^3} \sin^2 \vartheta \quad (7c)$$

where  $Q_0^2$  is the amplitude of the induced electric dipole with frequency  $\omega_s$ .

By use of Eqs. 5 to 7, the intensity of the scattered light per unit solid angle is now given by

$$I(t) = AE_0^2 [k_0^2 \cos^2 \omega_L t + k_1^2 \cos^2 (\omega_L - \omega_s)t + k_2^2 \cos^2 (\omega_L + \omega_s)t] + \text{cross terms}, \quad (8)$$

where

$$k_0^2 = \alpha_0^2 \omega_L^4, \quad (9)$$

$$k_1^2 = \frac{1}{4} \left( \frac{\partial \alpha}{\partial Q} \right)_0^2 Q_0^2 (\omega_L - \omega_s)^4, \quad (10)$$

$$k_2^2 = \frac{1}{4} \left( \frac{\partial \alpha}{\partial Q} \right)_0^2 Q_0^2 (\omega_L + \omega_s)^4. \quad (11)$$

The cross terms in Eq. 8 can be neglected since the power they radiate averages to zero over time intervals of sufficient length. This can easily be verified by calculating the time average

$$I = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau I(t) dt = \frac{1}{2} AE_0^2 (k_0^2 + k_1^2 + k_2^2). \quad (12)$$

From Eq. 8, we expect that the scattered light will have peaks at the frequencies  $\omega_L$  and  $\omega_L \pm \omega_s$ . This can directly be verified by calculating the frequency dependence of the scattered light, the power spectrum, which is obtained by forming the square of the Fourier transform of  $M(t)$ , namely

$$P(\omega) = A \lim_{\tau \rightarrow \infty} \frac{2}{\tau} \left| \int_{-\tau/2}^{\tau/2} \ddot{M}(t) \exp(-i\omega t) dt \right|^2 \quad (\text{Brüesch 1986}).$$

The classical theory thus correctly predicts the occurrence of the Stokes and anti-Stokes lines, but leads to an incorrect ratio of their intensities. From Eqs. 10 to 12, it follows that the ratio of the intensities of the Stokes and anti-Stokes lines should be

$$\frac{I_{\text{Stokes}}}{I_{\text{anti-Stokes}}} = \frac{(\omega_L - \omega_s)^4}{(\omega_L + \omega_s)^4}, \quad (13)$$

which will certainly be less than unity, whereas experimentally it is found that the Stokes lines are more intense than the anti-Stokes lines. This inconsistency is eliminated in the quantum theory of the Raman effect.

At the beginning of this section “A Simple Classical Approach”, it was pointed out that, in general, the direction of the induced dipole moment  $\mathbf{M}$  does not coincide with the direction of the electric field  $\mathbf{E}$ . In most cases, the polarizability within a molecule will be greater in one direction than in another, so that the induced dipole will tend to align itself with the direction of greatest polarizability in the molecule rather than with  $\mathbf{E}$ . In such a case, Eq. 1 (neglecting the quadratic term) has the form:

$$\begin{aligned} M_x &= \alpha_{xx}E_x + \alpha_{xy}E_y + \alpha_{xz}E_z \\ M_y &= \alpha_{yx}E_x + \alpha_{yy}E_y + \alpha_{yz}E_z \\ M_z &= \alpha_{zx}E_x + \alpha_{zy}E_y + \alpha_{zz}E_z \end{aligned} \quad (14)$$

Considering the energy of the polarized system, it may be shown that  $\alpha$  is a symmetrical tensor with:

$$\alpha_{xy} = \alpha_{yx}, \alpha_{xz} = \alpha_{zx} \text{ and } \alpha_{yz} = \alpha_{zy}$$

and the  $\alpha_{ij}$  components of the polarizability tensor can be described by a real symmetric matrix.

It can also be shown that there exists always a coordinate system with axes  $(x', y', z')$  such that the relation between  $\mathbf{M}$  and  $\mathbf{E}$ , when referred to these axes, assumes the simple form

$$\begin{pmatrix} M'_{x'} \\ M'_{y'} \\ M'_{z'} \end{pmatrix} = \begin{pmatrix} \alpha'_{x'x'} & 0 & 0 \\ 0 & \alpha'_{y'y'} & 0 \\ 0 & 0 & \alpha'_{z'z'} \end{pmatrix} = \begin{pmatrix} E'_{x'} \\ E'_{y'} \\ E'_{z'} \end{pmatrix} \text{ or } \mathbf{M}' = \alpha' \mathbf{E}', \quad (15)$$

where  $\alpha'$  is a diagonal matrix. Such axes are called principal axes of polarizability and are easy to find for a symmetrical system (molecule or crystal), since they must coincide with the symmetry axes present and be perpendicular to any plane of symmetry.

It is convenient to define the system  $(x, y, z)$  in such a way that it coincides with the principal axes of the equilibrium configuration of the molecule or crystal. For the equilibrium configuration, the components of  $\alpha$  are

$$\alpha_{ij} = \delta_{ij}\alpha_{ij}^{(0)} \quad (16a)$$

or written differently

$$\alpha = \begin{pmatrix} \alpha_{xx}^{(0)} & 0 & 0 \\ 0 & \alpha_{yy}^{(0)} & 0 \\ 0 & 0 & \alpha_{zz}^{(0)} \end{pmatrix}. \quad (16b)$$

If as a result of thermal fluctuation the system is in a distorted configuration, there will be a new principal axes system, the system  $(x', y', z')$ , which in general will not coincide with  $(x, y, z)$ . The polarizability will again be diagonal if referred to  $(x', y', z')$  but not necessarily if referred to  $(x, y, z)$ . In the latter case, each component  $\alpha_{ij}$  can be expanded in terms of the normal coordinates  $Q_s$  (as for Eq. 2) to give

$$\alpha_{ij} = \alpha_{ij}^{(0)} + \sum_s \alpha_{ij,s} Q_s + \frac{1}{2} \sum_{s' s''} \alpha_{ij,s''} Q_{s'} Q_{s''} + \dots \quad (17)$$

For a given normal coordinate  $Q_s$ , we may define changes in the polarizability components

$$\Delta \alpha_{ij,s} = \alpha_{ij,s} Q_s = \left( \frac{\partial \alpha_{ij}}{\partial Q_s} \right)_0 Q_s \quad (18)$$

and a matrix with elements

$$\alpha_{ij,s} = \left( \frac{\partial \alpha_{ij}}{\partial Q_s} \right)_0, \quad (19)$$

namely

$$\delta \alpha^{(s)} = \begin{pmatrix} \alpha_{xx,s} & \alpha_{xy,s} & \alpha_{xz,s} \\ \alpha_{yx,s} & \alpha_{yy,s} & \alpha_{yz,s} \\ \alpha_{zx,s} & \alpha_{zy,s} & \alpha_{zz,s} \end{pmatrix}. \quad (20)$$

Substituting Eq. 17 into Eq. 14 and using

$$E_j = E_{0j} \cos \omega_L t \quad (21a)$$

$$Q_s = Q_0 \cos \omega_s t, \quad (21b)$$

we obtain

$$M_i(t) = a_i \cos \omega_L t + \sum_s b_{si} [\cos(\omega_L - \omega_s)t + \cos(\omega_L + \omega_s)t] \quad (22)$$

where

$$a_i = \sum_j \alpha_{ij}^{(0)} E_{0j} \quad (23a)$$

$$b_{si} = \frac{1}{2} Q_{s0} \sum_j \alpha_{ij,s} E_{0j} \quad (23b)$$

Equation 22 is a generalization of the corresponding Eq. 5 for diatomic molecules. The first term in Eq. 22 is responsible for Rayleigh scattering while the other

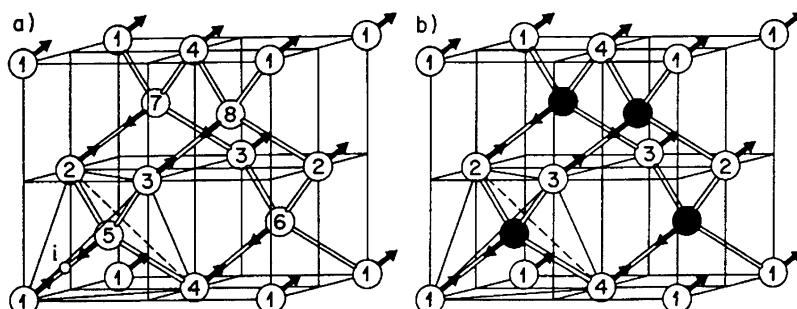
terms give rise to Raman scattering. We therefore expect Raman lines at the frequencies  $\omega_L \pm \omega_s$ . From Eq. 23b, it is evident that the normal mode  $Q_s$  will appear in the Raman spectrum only if at least one of the six components  $\alpha_{ij,s}$  of the matrix  $\delta\alpha^{(s)}$  is different from zero. If this is the case, the normal mode  $Q_s$  is Raman active. Whether or not a normal mode is Raman active depends on the symmetry of the equilibrium configuration and of the symmetry of the normal modes  $Q_s$ . This can be determined by consulting group character tables for the relevant point group. The tables are also useful to establish which crystal orientations for a single crystal will produce Raman activity for a given mode.

The scattering efficiency,  $I$ , depends on the polarization vector of the incident ( $e_i$ ) and scattered ( $e_s$ ) light and is given by

$$I = C \sum_j |e_i \cdot \mathbf{R}_j \cdot e_s|^2 \quad (24)$$

where  $C$  is a constant and  $\mathbf{R}_j$  is the Raman tensor of the phonon  $j$ .  $\mathbf{R}_j$  are obtained from group theoretical considerations (Loudon (1964), Cardona (1982)). They are second-rank tensors which are proportional to  $\alpha_{ij,s}$ , and they are used to calculate the polarization selection rules. Loudon (1964) derived the Raman tensors for each of the 32 crystal classes (symmetry point groups).

Figure 2 visualizes the vibrational modes that are Raman active for diamond and cubic boron nitride (cBN). The diamond structure (Fig. 2a) is composed of two fcc lattices displaced from each other by one-quarter of a body diagonal. The structure possesses a center of inversion symmetry  $i$  which is, however, not located at the atoms, but at the midpoint of each line connecting nearest-neighbor atoms. The triply degenerate  $\mathbf{q} = 0$  TO mode is shown by arrows and can be pictured as a relative motion of the two interpenetrating fcc lattices against each other. The TO mode resembles the normal oscillation of a homonuclear diatomic molecule and is Raman active but not infrared active.



**Fig. 2** In crystals with the diamond structure (a) and with the zinc blende structure (b), the TO modes at  $\mathbf{q} = 0$  (indicated by arrows) are Raman active (Brüesch 1986). Reprinted by permission from Springer Customer Service Centre GmbH: Springer, Phonons: Theory and Experiments II by Peter Brüesch, © 1986 (<https://doi.org/10.1007/978-3-642-52263-5>)

The cBN structure (Fig. 2b) does not have inversion symmetry and the arrows indicate the optical mode at  $q = 0$ ; it resembles the normal oscillation of a heteronuclear diatomic molecule and is thus expected to be Raman active. Since cBN has some polarity to the bonds, the macroscopic electric field associated with the vibration removes the degeneracy of the optical mode yielding a doubly degenerate TO mode and a LO mode.

### **The Width of the Raman Line**

In addition to information contained in the position of the Raman peak, the relative value of the full width at half maximum (FWHM) of the Raman peak is also of importance. The Raman linewidth in general can be broadened via several mechanisms. The main mechanisms applicable here are homogeneous (lifetime) broadening (Borer et al. 1971) and broadening due to the size effect of the crystal (Nemanich et al. 1981; Ager et al. 1991). The Raman linewidth of natural diamond is related to the lifetime of the phonon created in the Raman process. The primary broadening mechanism in a perfect crystal is decay of the optical phonon into two acoustic phonons with opposite wave vectors. However, in imperfect crystals, the phonon created in the Raman process can also decay (scatter) at grain boundaries and at defect sites, further reducing the lifetime and broadening the Raman line. The theory of the homogeneous spectral line shape predicts that the linewidth is inversely proportional to the phonon lifetime and that the line shape is symmetric and of a Lorentzian form. The FWHM of a Raman peak can thus be used as a convenient ranking indicator of degree of plastic deformation as the FWHM is sensitive to “crystal disorder” via the homogeneous broadening mechanism.

The other possible mechanism which results in Raman line broadening is phonon confinement in a small domain size. This well-established model, first developed to explain Raman peak shifts and line shapes in semiconductors such as Si and GaAs (Richter et al. 1981; Campbell and Fauchet 1986), is based on the uncertainty principle, which states that the smaller the domain size, the larger the range of different phonons (with different  $q$  vector and different energy) that are allowed to participate in the Raman process. Hence the broadening of the Raman line in this case is due to the spread in phonon energy. According to this model, the linewidth gets broader and the Raman peak shifts to a lower frequency and the line shape becomes asymmetric toward lower frequencies as the domain size decreases.

### **The Effect of Stress on the Raman Line**

In the previous sections, it was shown that the Raman effect is due to the interaction of photons with vibrational mode(s) (phonons) in a crystal. Phenomenologically, the effect of applying stress to a crystal that has Raman active vibrational modes may be explained as follows: If no changes are introduced to the unit cell of that crystal except to slightly change the frequency of the vibration, then this will result in a slight change in the position of the observed Raman bands. Such a change as imagined above can be introduced by way of an applied stress to the unit cell.

A slight compression will result in a slight change in the force constants of the vibration, thus leading to a change in the frequency of the vibration. The same argument can be applied if the unit cell dilates slightly due to the presence of tensile stresses in the lattice.

If the applied stress is isotropic, then no change takes place with respect to the Raman band except for a change in position since the symmetry of the unit cell is not changed. If uniaxial stress is applied, then the unit cell will deform preferentially in one direction and the symmetry of the cell changes locally. This leads to a directional dependence in position of the Raman band, i.e., the magnitude of the shift due to stress would depend on the direction in which it is applied and the orientation of the measurement relative to the direction of stress. If the Raman band is degenerate, then the application of uniaxial stress will lead to a lifting of the degeneracy and a consequent splitting of the Raman band.

### The Hydrostatic Piezo-Raman Coefficient $\alpha$

One of the first papers theoretically addressing the effect of stress on the Raman modes was that by Ganesan et al. (1970). The treatment according to Mossbrucker and Grotjohn (1997) and Anastassakis et al. (1970) is followed here, using diamond as an example. Diamond belongs to the crystal class m3m. Hence, in the absence of any strain, the  $\mathbf{k} = 0$  optical phonons in diamond are triply degenerate due to the cubic symmetry of the crystal. In the presence of strain, the differential equations for the  $\mathbf{k} = 0$  triply degenerate ( $F_{2g}$  symmetry) optical phonons in diamond type crystals to terms linear in the strain have the form

$$\bar{m}\ddot{x}_i = - \int_k D_{ik}x_k = - \left( D_{ii}^{(0)}x_i + \sum_{klm} \frac{\partial}{\partial \varepsilon_{lm}} D_{ik}\varepsilon_{lm}x_k \right), \quad (25)$$

where  $x_i$  is the  $i$ th component of the relative displacement of the two atoms in the unit cell,  $\bar{m}$  is the reduced mass of the two atoms, and  $D_{ii}^{(0)} = \bar{m}\omega_0^2$  is the effective spring constant of the  $F_{2g}$  modes in the absence of strain; the change in the spring constant due to applied stress  $\varepsilon_{lm}$  is defined by

$$\frac{\partial}{\partial \varepsilon_{lm}} D_{ik}\varepsilon_{lm} = D_{iklm}^{(1)}\varepsilon_{lm} = D_{ikml}^{(1)}\varepsilon_{ml} \quad (26)$$

and  $i, k, l$ , and  $m$  designate  $x, y$ , or  $z$ . Since diamond belongs to the cubic symmetry group, only three components of the tensor  $D^{(1)}$  are independent:

$$\begin{aligned} D_{1111}^{(1)} &= D_{2222}^{(1)} = D_{3333}^{(1)} = \bar{m}p \\ D_{1122}^{(1)} &= D_{2233}^{(1)} = D_{1133}^{(1)} = \bar{m}q \\ D_{1212}^{(1)} &= D_{2323}^{(1)} = D_{1313}^{(1)} = \bar{m}r \end{aligned} \quad (27)$$

From the dynamical Eq. (25) and assuming a harmonic solution, one obtains the following secular equation whose solutions yield the frequencies of the optical phonons in the presence of strain:

$$\begin{bmatrix} p\epsilon_{xx} + q(\epsilon_{yy} + \epsilon_{zz}) - \lambda & 2r\epsilon_{xy} & 2r\epsilon_{xz} \\ 2r\epsilon_{xy} & p\epsilon_{yy} + q(\epsilon_{xx} + \epsilon_{zz}) - \lambda & 2r\epsilon_{yz} \\ 2r\epsilon_{xz} & 2r\epsilon_{yz} & p\epsilon_{zz} + q(\epsilon_{xx} + \epsilon_{yy}) - \lambda \end{bmatrix} \equiv 0 \quad (28)$$

where  $\lambda = \Omega^2 - \omega_0^2$ , and  $\Omega = \omega_0 + \lambda/(2\omega_0)$  is the strain-dependent frequency of the optical phonons. The secular equation is referred to the system of crystallographic axes,  $x = [100]$ ,  $y = [010]$ , and  $z = [001]$ . Diagonalization of the secular matrix yields the set of three eigenvectors of the optical phonons in the presence of strain. The phonon deformation potentials  $p$ ,  $q$ , and  $r$  are related to changes in spring constants with applied strain and have been measured for many cubic materials, including diamond (Grimsditch et al. 1978).

For the hydrostatic case, a piezo-Raman coefficient may be calculated quite easily. Cubic diamond has a single first-order Raman line at  $\nu_0 = 1332.5 \text{ cm}^{-1}$  (room temperature), reflecting the frequency of the triply degenerate zone-center optical phonon. As mentioned, these phonon frequencies are sensitive to lattice deformations caused by mechanical stress. An arbitrary stress tensor can be generally decomposed into a hydrostatic component and a deviator, inducing a volume change of the unit cell and a distortion of the bond angle, respectively. The volume change of the unit cell due to the hydrostatic stress results in a linear shift of the triply degenerate Raman line. The Raman shift  $\Delta\nu_H$  relative to the stress-free state is related to the magnitude of the hydrostatic stress  $\sigma$  by (Mitra et al. 1969; Mohrbacher et al. 1996)

$$\Delta\nu_H = \frac{\nu_0\gamma}{B}\sigma. \quad (29)$$

Taking the room temperature values  $\gamma = 1.06$  for the diamond Grüneisen parameter (Grimsditch et al. 1978) and  $B = 442 \text{ GPa}$  for the bulk modulus (Field 1992), a hydrostatic piezo-Raman coefficient of  $-3.20 \text{ cm}^{-1}/\text{GPa}$  is obtained.

Experimentally, this linear dependence of the peak position of the Raman line on the applied stress was confirmed by Grimsditch et al. (1978). It was shown that upon applying compressive stress to diamond, the peak position of the Raman line shifts linearly with stress to a higher frequency. The dependence of the Raman shift  $\Delta\nu$  on the applied stress  $\sigma$  is

$$\Delta\nu = \nu - \nu_0 = -\alpha\sigma \quad (30)$$

where  $\alpha$  is the pressure or piezo-Raman coefficient and  $\nu_0$  is taken to be the Raman peak position of the natural diamond when no pressure is applied. Conversely, a

move of the Raman peak position to lower frequencies is associated with the presence of tensile stress in the diamond.

A similar discussion for silicon is given by De Wolf (1996) in a very good topical review.

The value of the coefficient  $\alpha$  for diamond has been calculated and measured in a diamond anvil cell (DAC) by several authors. Table 1 summarizes several of the reported values in the literature.

### Uniaxial Stress

The triple degeneracy of the zone-center optical phonon is lifted when deviatoric stress distorts the bond angles and removes the symmetry. Strains and changes in crystal symmetry are expected to alter the vibrational mode frequencies and intensities due to anharmonic effects and changes in selection rules. This causes the splitting of the single Raman line into either three singlets or into a singlet and a doublet. For uniaxial stress applied along either the [001] or the [111] directions, one finds that the threefold degeneracy of the  $\mathbf{k} = 0$  optical phonon is split into a singlet with the eigenvector parallel to the stress and a doublet with eigenvectors perpendicular to the stress. Uniaxial stress along the [110] direction is expected to completely remove the threefold degeneracy. There is also a shift in the frequency of the optical phonons to  $\Omega_0 = \omega_0 + \Delta\Omega_h$  due to the hydrostatic component of the applied stress (Anastassakis et al. 1970). It is found, however, that the centroid position of the split lines does not change under deviatoric stress. In particular cases (pure shear (traceless) stress), the splitting of the singlet is twice that of the doublet and opposite in sign (Grimsditch et al. 1978), whereas the doublet splits symmetrically in other cases. The centroid position of the Raman split line is thus entirely determined by the hydrostatic stress. Since the strain-induced changes in the Raman tensor are small, it is safe to assume that the scattering is determined by the first-order Raman scattering tensor. The relative magnitude of the shift for the two components is dependent on the direction of the stress. The Raman spectra of the

**Table 1** Values for the hydrostatic piezo-Raman coefficient  $\alpha$  of diamond reported in literature

Value of $\alpha$ ( $\text{cm}^{-1}/\text{GPa}$ )	Conditions of measurement	Reference
$2.8 \pm 0.3$	Hydrostatic (DAC)	Mitra et al. 1969
$2.87 \pm 0.1$	Hydrostatic (DAC)	Boppert et al. 1985
$2.96 \pm 0.11$	Hydrostatic (DAC)	Whalley et al. 1976
$3.20 \pm 0.2$	Hydrostatic	Grimsditch et al. 1978
3.1	Calculated	Van Camp et al. 1992
1.9	Reported average value for hydrostatic case	Bergman and Nemanich 1995
2.9	Reported as literature hydrostatic value	Ager et al. 1991
$2.88 \pm 0.17$	Average of published hydrostatic measurements in DAC	Catledge et al. 1996
$3.6 \pm 0.3$	Hydrostatic (tetrahedral cell)	Parsons 1977
$2.90 \pm 0.05$	Hydrostatic (DAC)	Hanfland et al. 1985

two components can be distinguished by polarization analysis. The linewidth of the two components does not change with increasing uniaxial stress; this is also true for the single line observed in hydrostatic stress experiments (Mitra et al. 1969; Anastassakis et al. 1970) (Table 2).

### Biaxial Stress

In the case of diamond on a substrate, e.g., CVD films on a silicon substrate or polycrystalline layers on a tungsten carbide backing, a so-called biaxial stress model has been invoked to explain the observed features of the Raman spectrum (Ager and Drory 1993; Catledge and Vohra 1995, Catledge et al. 1996, Nugent and Prawer 1998). In this model, it is assumed that the stress in the diamond film can be split into two components, namely an in-plane stress component parallel to the film or layer and a component normal to the film or layer.

For diamond, biaxial compressive stress in either {100} or {111} planes splits the Raman phonon into a singlet and a doublet (wave vector perpendicular and parallel to stress plane, respectively), where the doublet is at higher frequency in diamond (Ager and Drory 1993). Biaxial stresses along other axes lift the triple degeneracy completely.

In the case where a biaxial model is suspected to be applicable, but no splitting is observed, it may be that the peak splitting is too small to be resolved by the spectrometer. In this case, a “weighted” shift  $\Delta\nu = \gamma_3\Delta\nu_s + \gamma_3\Delta\nu_d$  can be used to calculate the stress ( $\sigma = -0.617 \text{ GPa/cm}^{-1} (\nu - \nu_0)$  for the unsplit peak) (Catledge and Vohra 1995) (Table 3).

**Table 2** Values for the piezo-Raman coefficient  $\alpha$  of diamond under uniaxial stress

Value of $\alpha$ ( $\text{cm}^{-1}/\text{GPa}$ )	Conditions of measurement	Reference
$2.2 \pm 0.2^{\text{a}}$	Uniaxial stress along [111]	Grimsditch et al. 1978
$0.73 \pm 0.1^{\text{a}}$	Uniaxial stress along [001]	
1.6 <sup>b</sup>	Calculated; strain along [111]	Van Camp et al. 1992
2.0 <sup>b</sup>	Calculated; strain along [110]	
1.6 <sup>b</sup>	Calculated; strain along [100]	
2.37	Value is for singlet; strain direction not specified	Sharma et al. 1985
1.04	Calculated; [110] compression; first peak	Gupta et al. 1989
1.97	Calculated; [110] compression; second peak	
0.67	Calculated; [110] compression; third peak	
1.68 <sup>c</sup>	[100] compression; singlet	Boteler and Gupta 1993
1.1 <sup>c</sup>	[100] compression; doublet	
0.93 <sup>c</sup>	[110] compression; first peak	Boteler and Gupta 1993
2.22 <sup>c</sup>	[110] compression; second peak	
0.09 <sup>c</sup>	[110] compression; third peak	

<sup>a</sup>Note that for these values, it is the relative splitting between the two peak components that is given and not the value of  $\alpha$  as defined in Eq. 30

<sup>b</sup>The authors do not state to which of the component(s) of the split peak these values apply

<sup>c</sup>These values were estimated from the published figures in this reference

**Table 3** Values for the piezo-Raman coefficient  $\alpha$  of diamond in the biaxial stress model

Value of $\alpha$ (cm <sup>-1</sup> /GPa)	Conditions of measurement	Reference
1.64	Calculated; for the [100] singlet	Ager and Drory 1993
2.37	Calculated; for the [100] doublet	
0.67	Calculated; for the [111] singlet	Ager and Drory 1993
2.86	Calculated; for the [111] doublet	
0.90	Calculated; for the [110] singlet	Ager and Drory 1993
2.25	Calculated; for the [110] doublet	
0.93	Calculated; polycrystalline diamond, singlet phonon	Ager and Drory 1993
2.60	Calculated; polycrystalline diamond, doublet phonon	
1.62	For unsplit peak; see text	Catledge and Vohra (1995)

Note that here the plane of the in-plane stress component of the biaxial stress distribution is specified rather than the direction in which the stress is applied, as with the uniaxial stress data

It is important to note that for a backscattering measurement geometry and for the case of a (100) facet, only the singlet would ever be observed (Ager and Drory 1993) and such a biaxial stress could never lead to splitting of the peak. If the stress is along some other direction or is uniaxial, this analysis no longer rigorously applies (Nugent and Prawer 1998).

It should be noted that there are conditions where at first glance a biaxial stress model may be relevant, such as cross-sections in 3D microelectronics components, but in fact the vertical stress component should not be neglected. This would, e.g., be for stress caused by through Si vias or microbumps, where the vertical stress component plays a large role in the measured Raman shift. The paper by De Wolf (2015) is particularly relevant here, discussing when it is necessary to take into account a triaxial stress model.

### The Effect of Temperature on the Raman Line Position

The lattice dynamics that give rise to the Raman effect are important in determining both thermal properties, such as specific heat and thermal conductivity, and high-temperature stability. Measurements of the temperature-dependent Raman line shifts and widths also allow comparison with predictions of anharmonic crystal theory.

The temperature dependence of the shifts and width can be understood in terms of an oscillator that reacts to the anharmonic potential within which the atoms move. The anharmonicity of the potential manifests itself in the temperature dependence of the Raman line shift, which can be written (in the absence of strain) as a result of two additive effects: the first is the change of vibrational frequencies due to thermal expansion (or volume change) and the second due to the change in the average vibrational energy at constant volume (a self-energy shift that arises from phonon coupling). The former affects all frequencies while the latter affects the higher-lying states more than the lower-lying states (Postmus et al. 1968). These additive effects can be written as (Herchen and Cappelli et al. 1993; Datchi and Canny 2004)

$$\Delta\nu = \left(\frac{\partial\nu}{\partial V}\right)_T \Delta V + \left(\frac{\partial\nu}{\partial T}\right)_V \Delta T \quad (31)$$

Any change in temperature will affect these two factors, thereby changing the position of the Raman peak as a function of temperature.

At higher temperatures, the anharmonicity of the potential within which atoms move will lead to increased phonon-phonon interactions that result in scattering and shorter lifetimes of the optical phonons with a concomitant increase in the population of acoustical phonons. Hence the Raman peak width in a single crystal is determined by phonon decay processes, and the width of a Raman line will increase with increasing temperature (Brüesch 1982). Elementary theoretical predictions of the change in TO Raman linewidth with temperature depend on the concept that at higher temperatures, the phonon occupation number  $n$  increases, leading to an increased probability of phonon decay. The process can be a TO-mode phonon decaying either into two longitudinal acoustic phonons of opposite momentum or into three acoustical phonons. If the phonon occupation number is defined as

$$n(\nu, T) = \frac{1}{[\exp(hc\nu/kT) - 1]}, \quad (32)$$

and the former decay mechanism is the sole decay mechanism, then for this two-phonon process, the predicted Raman linewidth  $\Gamma$  is (Hart et al. 1970)

$$\Gamma(T) = \Gamma(0) [1 + 2n(\nu/2, T)]. \quad (33)$$

$\Gamma(0)$  is an adjustable parameter that is fixed by fitting the curve through the lowest-temperature value available.

Equation 33 was extended by Balkanski et al. (1983) to include the contribution of three- and four-phonon processes, and for this model, the predicted Raman linewidth is

$$\Gamma(T) = A [1 + 2n(\nu/2, T)] + B [1 + 3n(\nu/3, T) + 3n^2(\nu/3, T)], \quad (34)$$

where A and B are constants.

## Instrumentation

In this discussion, the focus is on instrumentation for dispersive Raman spectroscopy, where a brief overview will be given of the components of instrumentation necessary to perform Raman spectroscopy and their respective roles. See also the handbook(s) referred to in the introductory paragraph of this chapter.

The following performance factors need to be considered when evaluating a Raman spectrometer: Raman photon collection efficiency, stray light rejection, spectral resolution, calibration stability, data acquisition speed, operational

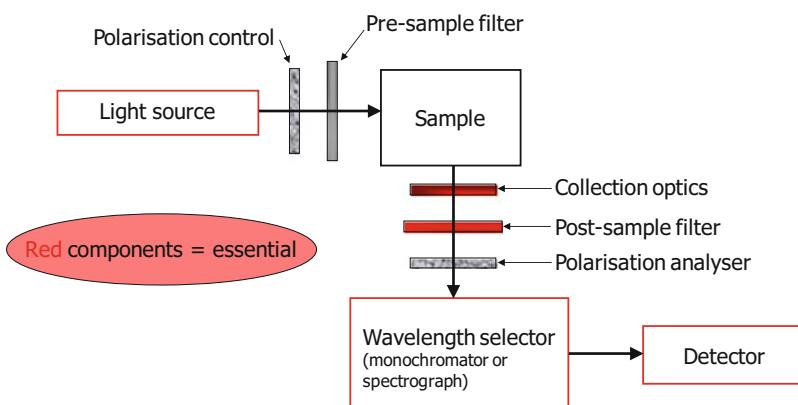
flexibility, and cost. As is frequently the case with scientific instrumentation, end-user requirements would dictate which of these factors are most important, as it is highly unlikely that one instrument would be optimal for all possible applications. Generally, instruments that can address a wide variety of problems due to the range of options fitted are significantly more expensive compared to models that are dedicated to only one type of analysis.

The basic components of an instrument are a light source, a sample, some form of light collection optics, a post-sample light filter, a wavelength selector, and a photon detector. Most of these components are under computer control (Fig. 3).

## Light Source

As can be deduced from the theory discussion, a high intensity monochromatic light source is required. The high intensity is needed to compensate for the inherent weakness of the Raman effect, and a monochromatic character is required to make sure there is only one Raman peak per phonon mode. The ideal light source is thus a laser, and Raman spectroscopy only really became viable as a measurement technique after the advent of the laser in the 1960s. Laser wavelengths from the UV to the near-IR are readily available. The workhorse laser used to be lasers based on ionized gas in a plasma tube (using argon, krypton, or helium-neon as the gas medium), but these are slowly being replaced by lasers based on diode technology, due to the lower power consumption and easier cooling requirements of diode lasers.

There are various reasons to consider using different excitation wavelengths. In section “[A Simple Quantum Mechanical Approach](#),” it was stated that the intensity of the Raman peak is proportional to the fourth power of the frequency (i.e., to  $\nu^4$  or  $1/\lambda^4$ ). As a general rule, a better response is obtained using blue or green excitation wavelengths, but that can lead to unwanted background fluorescence that often



**Fig. 3** Schematic layout of a Raman instrument, showing the components discussed in the text

swamps the relatively weak Raman peaks. A change of wavelength (often to red or NIR but also to UV) can help to avoid fluorescence problems.

A technique known as resonance Raman spectroscopy relies on the laser excitation wavelength being close enough to an absorption band in the sample material to significantly enhance the Raman signal.

A note is made here regarding the units used for the *x*-axis when displaying Raman spectra. The unit used is the wave number ( $\text{cm}^{-1}$ ), which is an energy unit. The laser excitation line is usually plotted at zero  $\text{cm}^{-1}$ , and the Raman peaks are then plotted as a Raman shift relative to the excitation wavelength. In this way, a specific Raman peak will have the same position on the *x*-axis independent of the excitation wavelength used.

The laser beam incident on the sample is usually passed through a narrow bandpass filter (pre-sample filter) to eliminate any possible plasma lines or other lines in the incident beam that may complicate the interpretation of the spectrum.

## Sample Considerations

Solid, liquid, or gas samples may be analyzed. Solid samples can range from single crystals to nano-sized powders, thin films, or really any object that can fit under the Raman microscope. Most modern Raman instruments focus the incident laser beam onto the sample either via a microscope attached to the Raman spectrometer, or a remote probe head that contains focusing and collection optics. The probe head is then connected to the Raman spectrograph via fiber-optic cables.

A flat sample surface is advantageous, but not critical in order to obtain good spectra. For uneven samples, the correct choice of microscope objective becomes important.

It is important to note here that as the incident laser beam is typically focused down to a spot of only a few microns in diameter, the power density ( $\text{W}/\text{cm}^2$ ) at the sample can reach  $\text{MW}/\text{cm}^2$  values fairly easily. It is thus relatively straightforward to locally heat the sample with the laser beam and even burn it, thus modifying the sample in the measurement process. As indicated in the theory section, the Raman peak position and peak width are functions of temperature; thus, it is important to be mindful of the laser power employed in order not to distort the results.

## Collection Optics

The scattered light can be acquired in a backscattering geometry, a  $90^\circ$  geometry, or a transmission geometry ( $0^\circ$  scattering). The backscattering configuration is most commonly employed. The backscattered light is collected via a microscope objective, and most micro-Raman instruments are equipped with a range of objectives. The important parameters to consider here are the numerical aperture (N.A.) and working distance of the objective. The N.A. is given by

$$\text{N.A.} = n \sin \theta \quad (35)$$

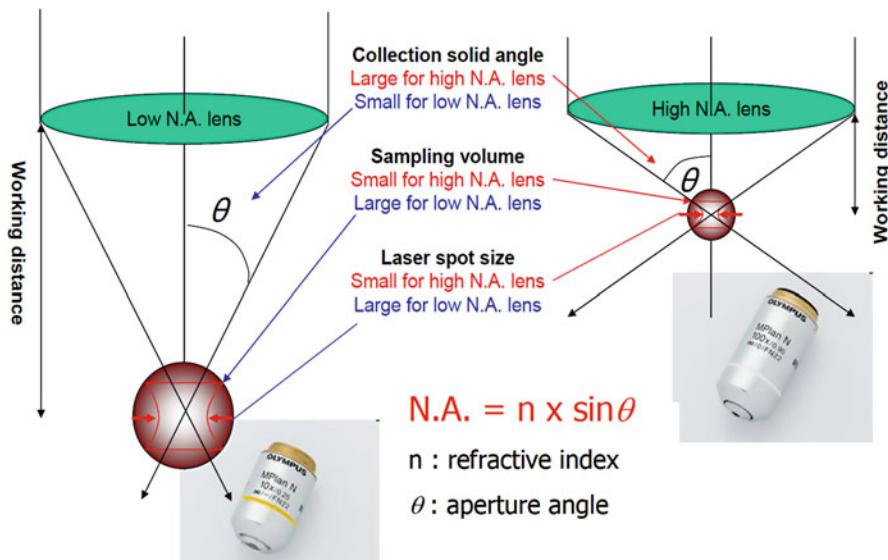
where  $n$  is the refractive index of the medium between the objective lens and the sample, and  $\theta$  is the half-angle of the collection cone, also known as the aperture angle (Fig. 4).

The diffraction limited diameter of the laser beam spot (Airy disk diameter) on the sample is given by

$$d = \frac{1.22\lambda}{\text{N.A.}}, \quad (36)$$

where  $\lambda$  is the wavelength of the light used and N.A. is as defined above. The smallest spot size, and thus the best lateral spatial resolution, is obtained using shorter wavelength light and an objective with a large N.A. Eq. 36 gives the ideal case. Actual spatial resolution will depend on the quality and alignment of the optics. Typically, these objectives have a small working distance, thus making it difficult to focus on very uneven samples. Special so-called long working distance objectives are available that give a larger working distance with only a small decrease in N.A.

Most modern micro-Raman instruments implement a confocal capability in the microscope in some way or other. The confocal capability allows for the axial spatial resolution (in the vertical or z-direction) to be changed by changing the size of the confocal pinhole. Decreasing the confocal pinhole diameter increases the axial resolution. The pinhole diameter typically varies from tens to hundreds of microns.



**Fig. 4** The influence of N.A. on the light collection properties of a microscope objective (Horiba Scientific)

A specific note is made here regarding the volume of sample “measured” by the laser beam. The diameter of the beam has been discussed earlier, but the transparency of the sample to the specific wavelength of the laser employed plays a big role in determining the sample volume. The “focal point” of the objective is in reality in the shape of an elongated cylinder, with axial length proportional to  $4\lambda / (\text{N.A.})^2$ . In a highly transparent material, the volume from which the Raman signal originates is directly related to the volume of the focal cylinder. If the sample material is only semitransparent to the excitation wavelength, then the penetration depth of the laser into the material, and thus the volume from which the Raman signal originates is strongly dependent on the absorption coefficient of the material at that wavelength. A good discussion in this regard is given by De Wolf (1996), and errors and artifacts in confocal Raman spectroscopy are discussed by, e.g., Everall (2010) and references therein. This paper and Everall (2014) are essential reading for researchers doing 2D and 3D mapping. The short note by Adar et al. (2010) contains some very useful practical advice.

## Post-Sample Filtering

The light scattered from the sample contains both Rayleigh scattered light and Raman scattered light. The Rayleigh scattered light is orders of magnitude more intense than the Raman scattered light and needs to be filtered out before the light reaches the detector. This can be done using double- or triple-stage monochromators, but in modern instruments, the filtering is done by notch or edge filters. These filters block transmission of the Rayleigh scattered light and either transmit both Stokes and anti-Stokes Raman photons (notch filters) or only Stokes Raman photons (edge filters). The latter is the standard filter employed in the majority of Raman instruments. Recent advances in filter technology have resulted in standard edge filters offering a cutoff wavelength between  $50 \text{ cm}^{-1}$  and  $80 \text{ cm}^{-1}$  (wavelength dependent), and special edge and notch filters offer cutoff positions as low as  $10 \text{ cm}^{-1}$ .

## Polarization Control

A detailed discussion on polarization control is not included here, but it is mentioned that a half-wave plate prior to the sample can be used to rotate the plane of polarization incident on the sample and a polarization analyzer post sample can be rotated to pass light in parallel or perpendicular configurations. This allows for very specific selection of certain Raman active modes and is usually confined to single crystal samples.

## Wavelength Selector

The wavelength selector, a spectrograph in most modern Raman instruments, is a device that disperses the frequencies contained in the light from the sample into their

constituent components in order to be able to detect the discrete Raman bands. For this function it makes use of a diffraction grating. The important parameter that characterizes a spectrograph is the spectral resolution that it can achieve, i.e., its ability to resolve closely spaced peaks. The spectral resolution is a function of the dispersion of the spectrograph, the entrance slit width, and the pixel size on the charge-coupled device (CCD) detector. In its turn, dispersion is a function of the focal length of the spectrograph, the groove density of the grating, and the excitation wavelength. In general, a long focal length spectrograph and a high groove density grating give the best spectral resolution. A high-quality spectrograph typically achieves a spectral resolution of  $1\text{ cm}^{-1}$  or better, but a resolution of  $3\text{--}6\text{ cm}^{-1}$  is adequate for most applications.

A narrow entrance slit improves spectral resolution, but at the cost of light throughput, so usually a compromise is made between signal intensity and spectral resolution when it comes to choice of entrance slit width.

A CCD detector does not detect a continuous energy spectrum, but consists of an array of pixels (typically 1024 or 2048 pixels). The dispersion of the spectrograph then determines the wavelength range dispersed over the array of 1024 (or 2048) pixels, so that a larger number of pixels usually results in a better spectral resolution.

## Detector

In modern instruments, the signal photons are usually detected by a CCD detector. This is an example of a multichannel detector (i.e., it simultaneously detects a range of energies). Historically, single-channel detectors such as photomultiplier tubes (PMTs) were used in conjunction with scanning multistage monochromators, but this is a relatively slow method of detection.

CCD detectors are close to ideal for Raman spectroscopy. They are based on silicon, cover a spectral range of 200 nm to 1100 nm with high quantum efficiency, and can be cooled for improved dark count performance.

It is mentioned here that Raman spectra are often plagued by so-called cosmic ray events (CREs). These manifest as random, very sharp, narrow peaks, often only one or two pixels wide and are due to random background radiation caused by radioactive decay of elements or random cosmic ray events (cosmic rays incident on the upper atmosphere, giving rise to secondary and tertiary particle cascades that reach the CCD).

With modern instrumentation, it is relatively easy to obtain a Raman spectrum, but optimization requires patience and experience.

## Advantages and Disadvantages of Raman Spectroscopy

### Advantages

1. Raman spectroscopy provides characteristic spectra for qualitative identification of material composition and structure.

2. There is a relative lack of interference from overlapping peaks and, e.g., aqueous solutions do not present a problem (compared to, e.g., IR spectroscopy).
3. A spatial resolution of about  $1\mu\text{m}$  is readily attainable using micro-Raman spectroscopy, with a resolution in the 500 nm range possible with the correct choice of wavelength and objective. Resolution in the range of tens of nm is possible with techniques such as tip-enhanced Raman spectroscopy (TERS).
4. No sophisticated sample preparation is necessary.
5. It is a nondestructive, *in situ* technique, where variable temperature and variable pressure measurements are relatively easy to conduct. Care should be taken not to locally heat the sample with the laser beam as indicated earlier.
6. Spectra are acquired in a relatively fast acquisition time (milliseconds to minutes, depending on the sample). This allows for high spatial resolution 2D and 3D Raman mapping with tens of thousands of points being acquired in a few hours or less.

#### Disadvantages

1. The inherent weakness of the Raman effect makes it difficult to detect the signal, but with modern sources (laser) and advances in instrumentation, specifically detectors, this is being overcome.
2. A fluorescence background can be intense enough to swamp the Raman signal but can be partially suppressed by choice of laser excitation wavelength and/or software methods for background subtraction.
3. Not all compounds are Raman active, i.e., have Raman spectra, and therefore a complementary technique such as IR spectroscopy is needed.
4. It can only probe bulk properties of materials which are transparent to the incident laser wavelength; otherwise, it is a surface or near-surface technique.

---

## Typical Applications

As indicated earlier, this discussion focuses on the use of Raman spectroscopy for phase identification and determination of stress and strain in inorganic materials. Raman spectroscopy has had wide application in the nondestructive evaluation of a broad range of materials, including semiconductor materials such as Si (see, e.g., Starman et al. (2003), Srikar et al. (2003), Beechem et al. (2007), and De Wolf (1996; De Wolf et al. 1996; De Wolf and Anastassakis 1999; De Wolf 2003, 2015)) and GaAs (e.g., Jothilakshmi et al. (2011)), as well as a range of films and composites (Colomban 2002; Colomban et al. 2006; de la Vega et al. 2011; Imanaka et al. 2009; Zhao and Wagner 2004). The paper by Colomban (2002) is a very useful review of the use of Raman spectroscopy to analyze strain and stress in fiber-reinforced ceramics and metal-matrix composites. In terms of hard materials, Raman spectroscopy has been used to map residual stresses in  $\beta\text{-Si}_3\text{N}_4$  (Muraki et al. (1997)) and used to characterize micro-stresses and defects in various polymorphs of SiC (Lee et al. (2004)).

This section also highlights the application of Raman to ultra-hard materials, where research on the NDT of single crystal diamond (Erasmus et al. (2011a) and polycrystalline diamond (PCD) tool-bits (Erasmus et al. (2011b) and Vhareta et al. (2014)) as well as other results are used to illustrate the capabilities and advantages of this technique. McNamara et al. (2015) have also applied Raman spectroscopy to studies of residual stresses in PCD.

## Determination of Stress

A first requirement for a study of local mechanical stress with micro-Raman spectroscopy is that the material exhibits Raman active modes, i.e., there are well-defined peak(s) in the Raman spectrum. Second, a reference sample that is considered stress free should be measured on the same instrument as where the samples with stress are to be measured. This gives a well-characterized reference position relative to which the stress-induced shift of the Raman peak(s) can be compared. The instrument should be carefully calibrated prior to the measurement and the calibration should be stable, i.e., it should not drift significantly over the period of the measurements. Either laser plasma lines, or emission lines from external sources such as a Hg discharge lamp or a Ne discharge lamp can be used as calibration references. It is also important to use an appropriate laser power to prevent local heating of the sample. As mentioned previously, the sample temperature will influence the peak position and peak width, and can therefore complicate the measurement of stresses, especially those of small magnitude.

Recapping, in an ideal crystal, the sample would be stress free and the Raman peak would be measured at its theoretical position. As soon as the crystal is subjected to stresses, be it externally imposed or due to internal imperfections, the frequency of the lattice phonons is subtly affected and thus also the frequency of the Raman peak. In this way, careful measurements of the shift of the Raman peak relative to its unstressed frequency can yield information on the stresses present in the sample. Compressive stress results in a shift of the peak to higher frequencies, and tensile stress to lower frequencies. In order to convert a degree of shift to a stress quantity, use is made of the coefficient of stress-induced shift as follows:

$$\nu_r - \nu_s = \alpha\sigma, \quad (37)$$

where  $\nu_r$  is the position of the stress-free Raman line,  $\nu_s$  is the measured frequency of the Raman line,  $\sigma$  is the stress at the point in question (in GPa), and  $\alpha$  is the coefficient of stress-induced frequency shift. For example, for diamond, the value of  $\alpha$  can be  $1.9 \text{ cm}^{-1}/\text{GPa}$  [assumptions leading to this value discussed in Bergman and Nemanich (1995) and Mohrbacher et al. (1996)]. Thus in order to measure a stress of 100 MPa, a change in wave number position of  $0.2 \text{ cm}^{-1}$  must be measurable. This requires instrumentation with very good spectral resolution. For larger stresses, there are less stringent constraints on the spectral resolution requirements. De Wolf (1996) discusses these constraints with respect to silicon.

The Raman linewidth broadening for the materials under consideration here is mainly attributable to homogeneous (lifetime) broadening. In this case, the full width at half maximum (FWHM) of the diamond Raman peak is inversely proportional to the phonon lifetime. Since the phonon lifetime is shortened by scatter at grain boundaries and defect sites, the FWHM can be used as a convenient ranking indicator of the degree of “crystal disorder” and hence as an indicator of the degree of plastic deformation. It is noted here that it is more difficult to measure small magnitude stresses for broader peaks, i.e., if the material has undergone significant plastic deformation or if the particle size is small enough for phonon confinement effects to broaden and shift the peak positions. In this latter case, it is even more important to have a reference material that is considered stress free.

### Impressions in Single Crystal Diamond: 3D Raman Mapping

The following material is discussed in more detail in Erasmus et al. (2011a).

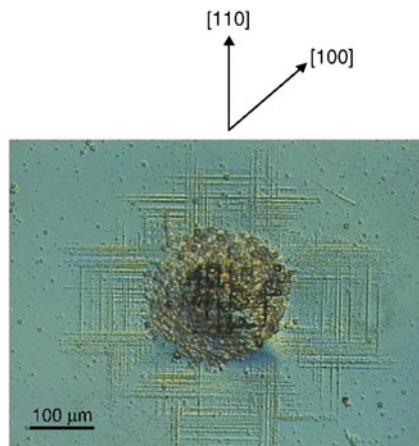
For the present purposes, an impression is defined as a plastic deformation made in diamond at high temperature where no brittle fracture has occurred, while an indentation is considered to be a deformation where brittle fracture has taken place. Impressions were made using an  $\text{Si}_3\text{N}_4$  impressor on a {001} face of a synthetic high pressure high temperature (HPHT) type Ib diamond with a mean contact pressure of 3.1 GPa within a vacuum furnace held at a temperature of 1400 °C. Nitrogen is present in the diamond as a substitutional impurity at the 500 ppm level. The diamond slip system is {111} <110>; applied stress is thus resolved onto {111} planes in <110> directions (Brookes 1992).

Raman spectra were acquired at ambient pressure and temperature. The excitation source was the 514.5 nm line of an argon ion laser, and the backscattered light was dispersed onto the liquid nitrogen-cooled CCD detector by a 1800 grooves/mm grating. The confocal pinhole was selected to give a specific Z-depth sampling thickness (exact details are given later in this section). A computer-controlled motorized X-Y stage in conjunction with manual Z-movement was used to map the spectral features within the three-dimensional volume of the impression. The position and width of the diamond Raman peak was extracted by fitting a Lorentzian function to the data using ORIGIN software (Fig. 5).

Figure 5 is a typical optical micrograph of an impression on single crystal synthetic diamond. The circular dark center corresponds to the area of contact of the impressor, and the cross-shaped pattern visible around its extremities is the so-called rosette (Brookes 1992). The fine horizontal and vertical lines are dislocations which have run out on the surface during the impression process, and have been etched to make them visible. It can be seen by referring to the indicators of crystallographic orientation at the top of Fig. 5 that the rosette shape is determined by the crystallographic plane used for the indentation and the slip system of diamond.

3D sections were obtained by selecting (100) and (110) crystallographic planes within the diamond and focusing the laser beam at appropriately chosen depths from the diamond surface along these planes to create effectively vertical “slices” within the bulk sample. Changes in the Raman peak position and peak width were

**Fig. 5** Optical micrograph of an impression on a single crystal synthetic diamond. Reproduced from J. Appl. Phys. 109, 013527 (2011), with the permission of AIP Publishing. (<https://doi.org/10.1063/1.3531548>)

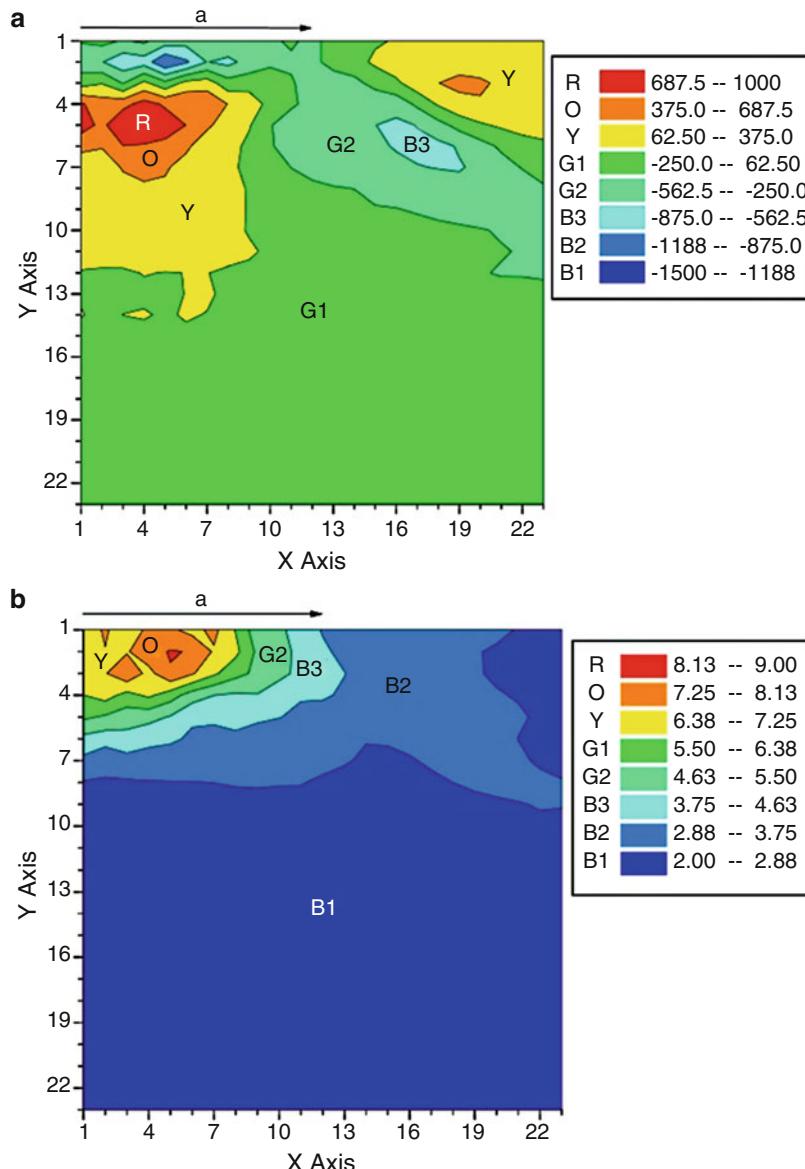


mapped along these sections. The center of the impression is at the origin of the arrow “a” in Figs 6 and 7. This arrow represents the radius of the circular impression on the surface of the diamond. Each map is thus of dimension  $2a$  on a side. The mapping grid size is  $23 \times 23$  points, with  $10 \mu\text{m}$  intervals. A  $50\times$  ultra-long working distance objective was used with a  $0.1 \text{ mm}$  pinhole, resulting in the laser beam focal barrel being  $10 \mu\text{m}$  and corresponding to the thickness of the diamond being sampled in the Z-direction for a particular focal position of the microscope.

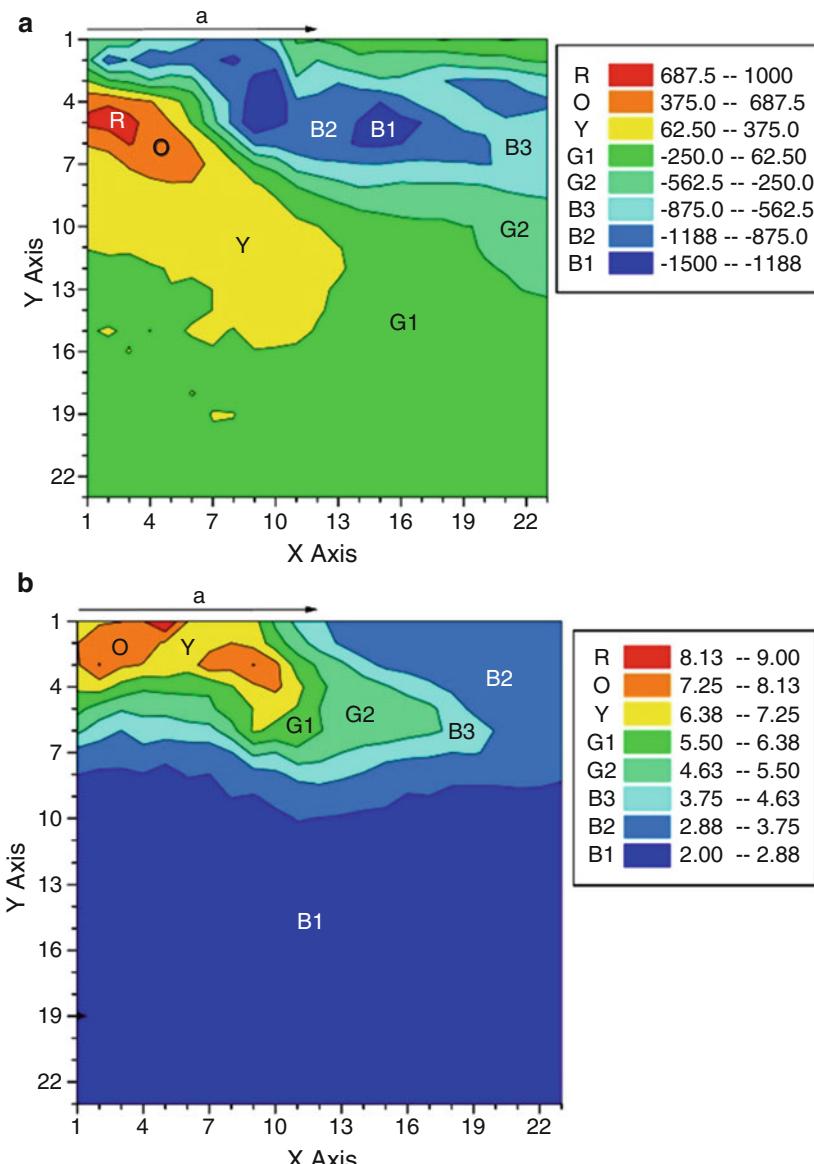
Compressive stress is indicated by negative stress values and its color range (green (minimum) to blue (maximum)), while tensile stress is indicated by positive stress values and its color range (yellow (minimum) to red (maximum)). The degree of deformation is depicted by a range for colors: the blue color corresponds to a “background” value for the FWHM of  $2.00\text{--}2.88 \text{ cm}^{-1}$  and the maximum peak width (red) corresponds to a peak width of  $8.13\text{--}9.00 \text{ cm}^{-1}$ .

The stress and degree of deformation maps for the (100) slice are shown in Fig. 6a, b, respectively, and are located between two arms of the rosette pattern associated with the impression. For the (100) slice, the largest compressive stress values and the greatest degree of plastic deformation are associated with the volume of diamond immediately below the contact area of the impressor. The greatest magnitude of tensile stress is also located immediately below the contact area of the impressor, but at a depth of approximately  $50 \mu\text{m}$ .

The maps for the (110) slice shown in Fig. 7a, b clearly illustrate the presence of the arm of the rosette in the [110] direction by the presence of an extended region of compressive stress. As shown in Fig. 7a, the maximum compressive stress in the arm ( $-1490 \text{ MPa}$ ) is associated with a volume located approximately  $50 \mu\text{m}$  below the surface of the diamond. The volume associated with the greatest magnitude of tensile stress is located immediately below the contact area of the impressor, also at a depth of approximately  $50 \mu\text{m}$ .



**Fig. 6** (a) Stress contour map of slice mapped into the diamond along the [100] direction. The center of the impression is at top left and the radius of the impression is “ $a = 110 \mu\text{m}$ .” The values in MPa of the respective stresses are given in the accompanying table. Negative values indicate compressive stress and positive values indicate tensile stress. (b) Contour map of the FWHM of the diamond Raman peak within the slice mapped into the diamond along the [100] direction. The values in  $\text{cm}^{-1}$  of the FWHM in each region are presented in the accompanying table. Reproduced from *J. Appl. Phys.* 109, 013527 (2011), with the permission of AIP Publishing. (<https://doi.org/10.1063/1.3531548>)



**Fig. 7** (a) Stress contour map of slice mapped into the diamond along the [110] direction. The center of the impression is at top left and the radius of the impression is “ $a = 110 \mu\text{m}$ .” The values in MPa of the respective stresses are given in the accompanying table. Negative values indicate compressive stress and positive values indicate tensile stress. (b) Contour map of the FWHM of the diamond Raman peak within the slice mapped into the diamond along the [110] direction. The values in  $\text{cm}^{-1}$  of the FWHM in each region are presented in the accompanying table. Reproduced from J. Appl. Phys. 109, 013527 (2011), with the permission of AIP Publishing. (<https://doi.org/10.1063/1.3531548>)

The degree of deformation mapped in Fig. 7b from the data on the FWHM of the Raman peak clearly shows that the maximum deformation is associated with the volume immediately below the contact area of the impressor.

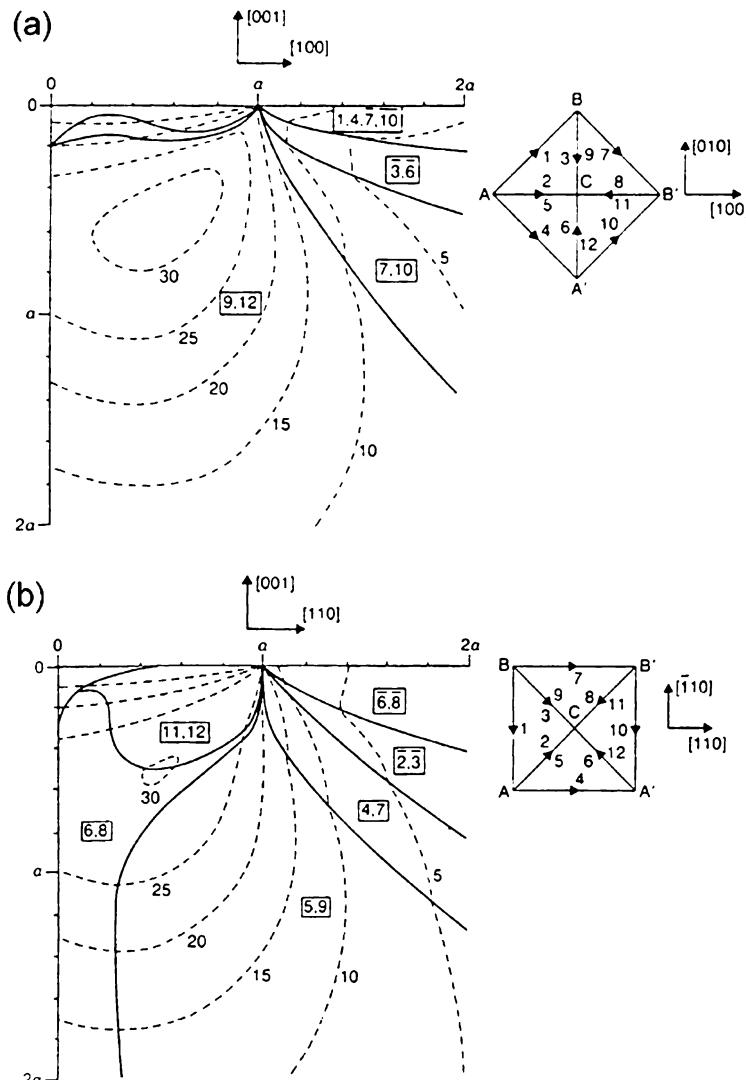
The respective magnitudes of the tensile stress measured for the (100) slice (maximum of 990 MPa) and the (110) slice (maximum of 830 MPa) slices agree well, and the respective degrees of deformation are also in good agreement (maximum widths of  $8.2 \text{ cm}^{-1}$  and  $9.0 \text{ cm}^{-1}$  for (100) and (110) slices).

It is also noted that whereas the degree of plastic deformation as indicated by the linewidth extends approximately 75  $\mu\text{m}$  from the surface of the diamond, the tensile stress field extends to approximately 150  $\mu\text{m}$  from the surface. It follows that the extent of the plastic deformation is less than that of the tensile stress field.

The most suitable model at present for indentation hardness is that suggested by Roberts (1988) who modeled the whole of the slip pattern over an area of  $2a \times 2a$ , where  $a$  is the contact area radius. This includes the areas directly beneath and adjacent to the contact point. Roberts used extended stress fields produced by a circular contact in an elastically isotropic material to predict the likely slip patterns and their associated stresses. This allowed complete 3D patterns of the relevant slip planes to be visualized together with the resolved shear stresses on them.

To model the magnitude of the resolved shear stresses which control plastic deformation in the bulk of a crystal, the crystallographic plane of the indented surface and the active slip systems must be identified. For this discussion, a (001) surface and  $\{111\} <110>$  slip systems are considered, the nomenclature indicating that the applied stress resulting in slip is resolved onto  $\{111\}$  planes in the  $<110>$  directions. Figure 8a, b (Brookes 1992; Brookes et al. 1990) are maps showing the highest resolved shear stresses and the slip systems which are associated with those stresses on a (100) plane and a (110) plane, respectively, where both of these planes represent sections which pass through the center of a circular contact area and are normal to the indented surface. For each of the (100) and (110) sections, the 12 possible slip systems in diamond cubic crystals are identified and numbered on the respective slip plane pyramids shown in Fig. 8a, b. In each of these pyramids, the respective apices “C” are assumed to be below the indented (001) plane. A further convention used here is that a bar over the slip system number indicates that material under that slip plane is sheared downward while a number without the bar corresponds to material above the slip plane being sheared downward, in other words, negative and positive slip, respectively. The actual slip systems subjected to the highest resolved shear stress in various regions beneath the contact area are shown in rectangular boxes between the solid lines, e.g., the numbers 9,12 in a box indicate that slip systems 9 and 12 are operative. The dashed lines indicate contours of constant resolved shear stress (RSS), and the number next to each contour indicates the percentage of  $P_m$  that is experienced along the contour, where  $P_m$  is the stress applied by the impressor over a circular contact area of radius  $a$ .

For Fig. 8a, i.e., on the (100) section or “slice,” the RSS values immediately below the surface are relatively low. The maximum shear stress is about  $0.316 P_m$  where this lies in a region where slip systems 9 and 12 are those most likely to be



**Fig. 8** (a) Modeled contours (broken lines) of resolved shear stress on the most stressed slip systems for contact on the (001) surface and for a section normal to the surface and along the [100] direction (Brookes et al. 1990; Brookes 1992). (b) Modeled contours (broken lines) of resolved shear stress on the most stressed slip systems for contact on the (001) surface and a section normal to the surface and along the [110] direction (Brookes et al. 1990; Brookes 1992). The values indicated are percentages of the stress  $P_m$  applied by the imprinter over a circular contact area of radius  $a$ . Solid lines divide regions where different slip systems are the most highly stressed. The slip system labeling is shown in the inserts, where the apices “C” of the pyramids of {111} slip planes (viewed from above the contact surface) are below the plane of the paper. A bar over the slip system number indicates that material under that slip plane is sheared downward (negative slip), while no bar corresponds to material above the slip plane being sheared downward (positive slip). Reproduced from J. Appl. Phys. 109, 013527 (2011), with the permission of AIP Publishing. (<https://doi.org/10.1063/1.3531548>)

activated. This region is  $\sim 0.58a$  beneath the center and  $\sim 0.54a$  from the center line of the impression. For Fig. 8b, i.e., on the (110) section or “slice,” the maximum shear stress is  $0.30 P_m$ , and this region is  $\sim 0.52a$  beneath the center of the impression and  $\sim 0.5a$  from the center line of the impression. Thus the zone of maximum resolved shear stress is roughly toroidal in shape (doughnut-shaped).

In Fig. 8a, b, the volume of maximum tensile stress corresponds fairly closely with the volume of diamond that would experience the maximum resolved shear stress during the impression process (broken line contours of 30% of the mean applied pressure  $P_m$ ). The volume of maximum tensile stress for the (100) slice also appears to be slightly larger than that for the (110) slice, which corresponds to the 30% contour being larger for the [100] slice. The maps in Figs. 6a and 7a thus show a close correspondence with the resolved shear stress contours in Fig. 8a, b, respectively.

From the description given above, the process resulting in the measured contours can be visualized. The first dislocation nucleation takes place in the volume of maximum resolved shear stress (30% broken line contour). As the impression process continues, dislocations multiply and move to fill the volume under impressor contact area. Material is thus “moving away” from the volume where the dislocations movement originates. At the end of the deformation process, this volume is thus “deficient” in material with the result that it is in tension. Although it would be expected that some relaxation of shear stress were to take place due to plastic deformation, the results show that most of the shear stress is “locked in” on cooling. A 0.316 fraction (maximum shear stress is  $0.316P_m$ , Fig. 8a) of  $3.07 \text{ GPa} = 970 \text{ MPa}$ , close to the measured maximum stress of 990 MPa for the (100) “slice” in the vicinity of the 30% contour. A 0.3 fraction (30% contour toroid in Fig. 8) of  $3.07 \text{ GPa} = 920 \text{ MPa}$ , comparable with the measured maximum for the (110) “slice” in an equivalent position of 830 MPa. The calculated resolved shear stress contours for the (100) “slice” encompass a larger volume than those for the (110) “slice” and this is reflected to a certain degree in the measured contours as well.

Related research in measuring residual stress in cubic boron nitride (cBN) may be found in Erasmus et al. (2000).

### Polycrystalline Diamond (PCD) Tools

The samples used in this work were commercial cylindrical drill-bits manufactured by Element Six (Pty) Ltd., of dimensions approximately 17 mm in length and 19 mm in diameter, consisting of a Co-WC substrate with a PCD layer sintered onto the circular face of the substrate. These drill-bits are used where very harsh working conditions prevail, such as rock cutting and drilling for oil. The diamond used was HPHT-synthesized synthetic grit (type Ib) with nitrogen as substitutional impurity present in excess of 500 ppm. Synthesis conditions for PCD generally involve applied pressures of 5 to 10 GPa, sintering temperatures of 1300 °C to 1800 °C

and sintering times of 5 to 60 minutes. The average grain size was 10  $\mu\text{m}$ . During the synthesis process, liquid phase sintering facilitated by molten Co takes place, resulting in an extended diamond structure with considerable direct diamond to diamond bonding and cobalt at the grain boundaries (see, e.g., Evans et al. (1984) and Uehara and Yamaya (1988)). The PCD layer was 0.6 to 1 mm thick, depending on the processing. The 514.5 nm line of an argon ion laser was used as the excitation wavelength. For the data reported here, samples were measured in the macro chamber of the Raman spectrometer, where the diameter of the spot size was  $\sim 300 \mu\text{m}$ . The PCD layer was nearly opaque, thus limiting the penetration depth of the laser to several tens of microns.

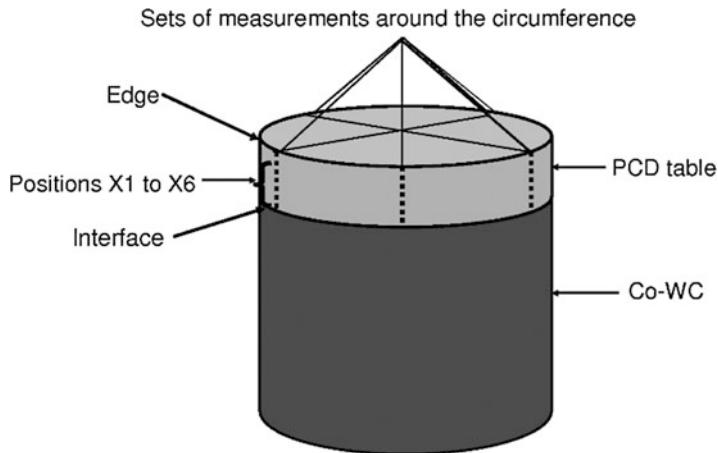
High temperature measurements were made *in situ* with a custom-built furnace, fitted into the macro chamber of the spectrometer. The furnace was mounted on an X-Y-Z adjustable stage with micrometer adjustments. Prior to heating, the furnace was evacuated to a high vacuum and backfilled with low-pressure argon gas to promote a uniform temperature of the sample. High temperature measurements were only performed on the samples with roughly lapped surfaces, as their spectra had the best signal-to-noise ratios. A single line of measurement points spaced at 2000  $\mu\text{m}$  intervals was used and the data averaged to give an average surface stress value for the PCD layer. The diamond Raman peak position was determined by fitting a Voigt function to the corresponding data.

### Effect of Annealing on PCD Properties

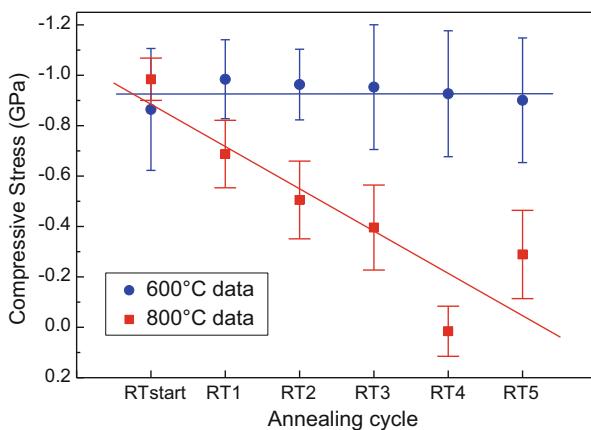
In order to investigate the effect that repeated cycling to a specific temperature would have on the tool-bit properties, several points ( $n = 10$  to 11) were measured across the surface of the PCD layer or table (Fig. 9), using the X-Y-Z micrometer-driven mounting stage of the furnace to move the sample. The surface stress state was measured at room temperature (21 °C) prior to annealing and after each annealing cycle for 5 cycles. During each heating cycle, the sample was heated at 15 °C/min to the set temperature, held at the set temperature for 30 minutes, and then cooled at 15 °C/min back to room temperature. In practice, the furnace cooled at the set rate to  $\sim 200$  °C, below which the cooling rate decreased substantially. This did not affect the results, as the temperatures where changes were expected to happen, were substantially above this temperature.

Results for five annealing cycles to 600 °C in an argon atmosphere are plotted in Fig. 10. It is clear that the surface compressive stress as measured at room temperature remains unchanged at approximately 0.9 GPa under these temperature and annealing conditions, indicating that the tool retains its properties. The data for the FWHM of the diamond Raman peak follows a similar trend, i.e., the FWHM remains unchanged at around 12.5  $\text{cm}^{-1}$  with cycling to 600 °C (Fig. 11).

Results for five annealing cycles to 800 °C in vacuum of a different tool-bit from the same manufacturing batch are also plotted in Fig. 10. Here it is clear that the annealing temperature and cycling are affecting the surface compressive stress, with



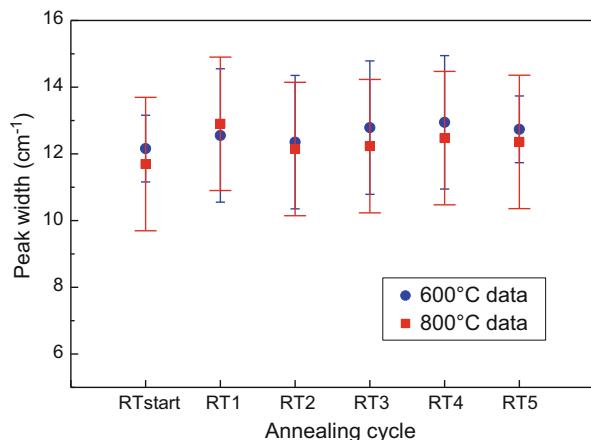
**Fig. 9** Diagram of the PCD tool-bits, showing the measurements on the PCD with X1 near the edge and X6 as the point near the interface region



**Fig. 10** Surface stress state for the PCD layer of a tool-bit as measured at room temperature after each of 5 cycles to 600 °C and 5 cycles to 800 °C. The solid lines are linear fits to the data to illustrate the trends. The data in the figure represent the average stress value across the surface, and the error bars represent one standard deviation in the data. Reproduced from Diam Relat Mater 20 (2011) 907–911, with the permission of Elsevier B.V. (<https://doi.org/10.1016/j.diamond.2011.03.018>)

the magnitude of the stress decreasing from 0.9 GPa to considerably lower values with increasing cycling. The tool-bit properties are thus being modified permanently. The FWHM in this case remains unchanged at around  $12.2\text{ cm}^{-1}$  with cycling (Fig. 11). Although substantially larger than the FWHM for single crystal diamond, these values are consistent with those reported by other authors (Ager et al. 1991;

**Fig. 11** FWHM of the PCD layer of a tool-bit as measured at room temperature after each of 5 cycles to 600 °C and 5 cycles to 800 °C. The data in the figure represent the average width value across the surface, and the error bars represent one standard deviation in the data. Reproduced from Diam Relat Mater 20 (2011) 907–911, with the permission of Elsevier B.V. (<https://doi.org/10.1016/j.diamond.2011.03.018>)



Bergman and Nemanich 1995) for polycrystalline diamond. The large width is ascribed to both the micro-defects that form due to fracture during the pressure ramp of the synthesis cycle, as well as the plastic deformation that takes place during sintering.

A surface compressive stress state in the PCD layer is desirable, as it would act to prevent the formation and slow the propagation of cracks in application, thus effectively acting as a toughening mechanism. It is clear from our results that the higher annealing temperature thus has a negative effect on the tool-bit properties.

Several possible mechanisms exist that can account for the observed trends in surface stress state with annealing temperature. The temperatures for cyclic annealing were chosen to assist in discriminating between mechanisms. Cobalt undergoes a phase transition from hcp to fcc around 420 °C. It is thus anticipated that annealing at 600 °C would indicate any effect that this process has on the surface stress state. The data shows that the surface stress state remains essentially unchanged during annealing to 600 °C, showing that the cobalt phase transition does not affect the surface stress state. Cobalt-catalyzed graphitization of diamond has an onset temperature of around 750 °C at ambient pressure, and thus annealing to 800 °C would show if this mechanism affects the surface stress state. Our data suggests strongly that this mechanism is mainly responsible for the change in surface stress state. It is noted that PCD heated to 1000 °C shows clear evidence of graphitic carbon in the Raman spectra measured on the surface of the PCD.

The results for the FWHM data indicate that no annealing of the diamond material in the PCD layer is taking place, as the FWHM is unchanged with repeated cycling for both temperatures. This suggests that the stress relieving mechanism is not due to a change in the diamond grains, e.g., possible dislocation movement, but due to changes in the diamond-matrix boundaries. This supports the role of a mechanism such as the cobalt-catalyzed graphitization mentioned in the previous paragraph.

The constituent diamond for the PCD layer in the tool-bits is type Ib. It has been shown that the brittle-ductile transition temperature (BDTT) for single crystal type Ib diamond is 750 °C (Brookes et al. 1990) suggesting that this might influence the peak width data in Fig. 11. Due to limited thermal stability of diamond-cobalt materials around 800 °C, flow stress data for these materials are inconclusive, but flow stress studies on diamond/β-SiC aggregates have shown that in the range of 800 °C to 1250 °C, time-dependent plastic deformation does not occur (Harris et al. 2001). This indicates that for polycrystalline diamond materials, the BDTT is at a higher value than for single crystal diamond. The absence of a change in the Raman peak width data with repeated cycling thus suggests that the BDTT has not been reached yet for the tool-bits considered in this study. Processes in the cobalt matrix and/or graphitization of the diamond are thus the most likely mechanisms to account for the observed reduction in compressive stress.

At room temperature, the overall stress is compressive around the circumference of the tool-bit (Fig. 9) and is largest at the PCD/WC interface region and lower at the edge confirming the predictions of finite element modeling (see also next section). Compressive stress in the PCD layer is desirable as it opposes the formation and propagation of cracks.

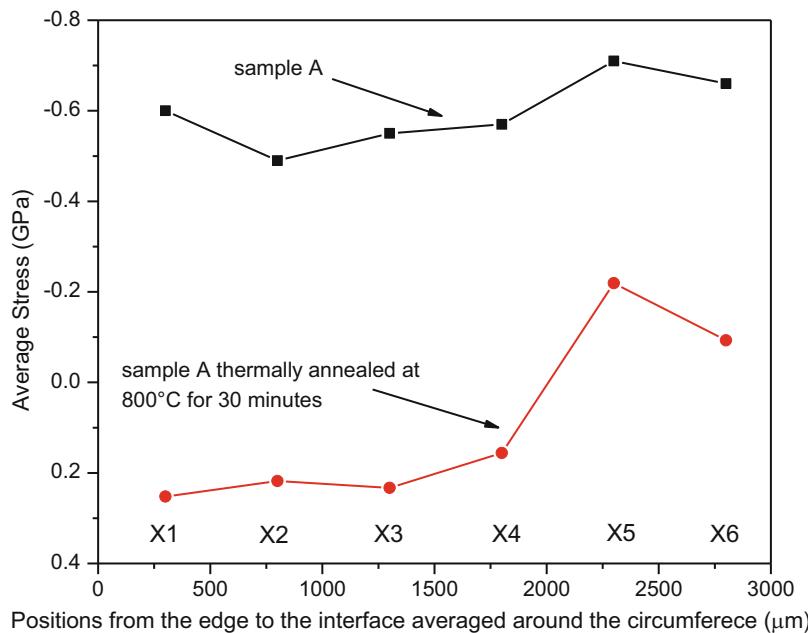
There was a significant change in the PCD stress magnitude after annealing. This implies a negative effect of annealing at 800 °C on the tool-bit properties, as crack initiation and propagation are more likely in the presence of reduced compressive stress or any tensile stress. This is a significant observation in determining maximum temperature conditions under which the tool-bit may be operated. 2D Raman mapping of the top of the PCD table after annealing revealed no sign of graphitization of the PCD. Only after a further annealing at 900 °C was any evidence of graphite noted in the region of the interface.

### Stress Gradients in the PCD Layer

In order to investigate the possible stress gradients that are present in the PCD table, Raman spectroscopy was used to measure a series of points between the edge of the PCD table and the PCD-Co/WC interface.

The tool-bits were measured in the macro chamber of the spectrograph in a backscattering configuration with a laser beam diameter of ~100 μm. Data points were measured from a point 300 μm from the top edge of the PCD table (point X1) at 500 μm intervals to the PCD/WC interface (point X6), as illustrated in Fig. 9. The surface stress state was measured at room temperature (23 °C) in all cases. Subsequent annealing of the tool-bits took place at 800 °C for 30 minutes in a Carbolite tube furnace under UHP argon. Peak positions were extracted from the data by fitting a linear combination of Lorentzian and Gaussian line shapes. The data in Fig. 12 have been averaged around the circumference of the tool-bit to show the variation of stress between the edge of the PCD table and PCD/WC interface.

The stress distribution between the edge of the PCD table and PCD/WC interface shows a trend, with a smaller compressive stress at the edge and a larger compressive stress near the interface. Measurements carried out after the high temperature annealing (800 °C for 30 minutes) show a significant reduction of stresses in the



**Fig. 12** Effect of temperature on the PCD stress for sample A before and after annealing to 800 °C for 30 min

tool-bit. The residual stresses are now largely tensile in the top section of the PCD layer (i.e., nearest the operating face) and are still compressive near the PCD/WC interface. The stress distribution observed prior to annealing has been amplified by the annealing process.

### Fatigue Processes in PCD

Polycrystalline diamond (PCD) cylindrical tool-bits used in oil well drilling are susceptible to fracture due to the hostile environment of randomly occurring impact loads to which they are subjected. The fact that the tool-bits fail after repeated use suggests the possibility of fatigue type processes in PCD. The study of stress fields on the surface of the PCD thus becomes crucial in the quest to have extended lives for these tool-bits. Some of the results from Vhareta et al. (2014) are given here in some detail.

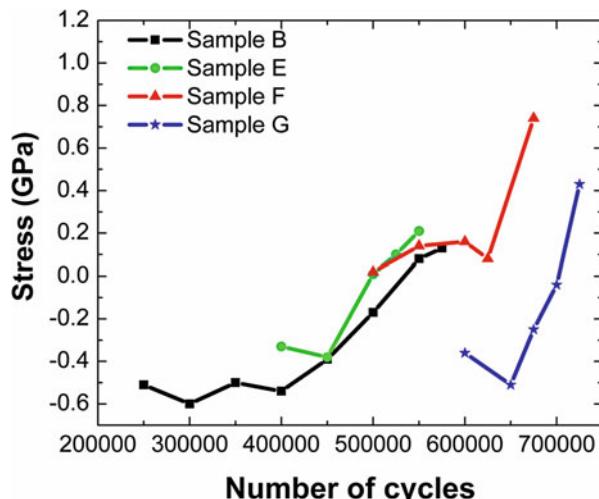
Since the diamond Raman peak reveals both the nature and magnitude of the stress present in the material, this technique can be employed as a nondestructive measurement tool to investigate these stress fields. Raman stress measurements at room temperature were carried out using a 36-point mapping array in area close to the size of the PCD samples. The PCD layers were detached from their respective substrates using electron discharge machining (EDM) and slightly polished. The samples were thus in the shape of disks. The fatigue cycles were undertaken using a ball on three ball fatigue jig. The mapping points provided histograms of the

magnitude and nature of these small individually stressed regions showing a general compressive stress for the lower numbers of fatigue cycles which deteriorates to a high proportion of tensile regions with increasing number of fatigue cycles. The data are also illustrated by 2D surface maps as an alternative mode of presentation, again confirming the change from surface stresses being dominantly compressive to dominantly tensile with exposure to the higher numbers of fatigue cycles. Whereas a general compressive stress is desirable in the PCD layer as it inhibits the propagation of cracks, on the contrary, tensile stresses facilitate the formation of cracks ultimately leading to catastrophic failure of the tool-bits. Only selected data is given here; the histograms and 2D maps can be found in Vhareta et al. (2014).

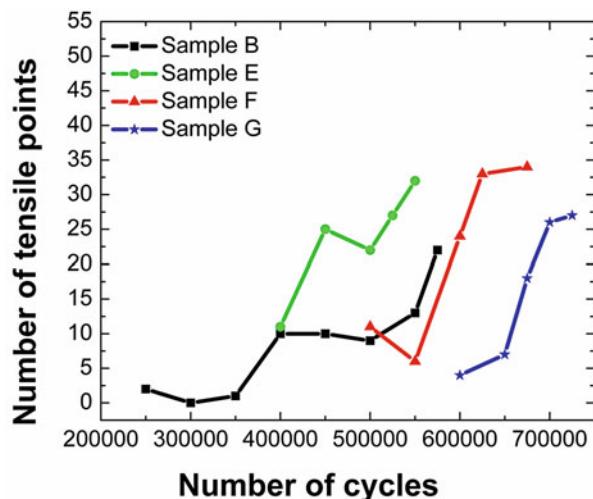
For comparison purposes, Figs. 13 and 14 illustrate the respective effects of the multiple fatigue cycles for four samples, namely, B, E, F, and G. The average stress variations for the samples are shown in Fig. 13, while in Fig. 14, the number of mapping points exhibiting tensile stress is indicated. Although there are obvious variations in detail in the results for the various samples, there are also common trends. In particular, the average surface stress is compressive during the earlier phase in which the number of fatigue cycles remains relatively limited while in the later phase of a large number of cycles, the average surface stress becomes tensile in nature for all the samples. A comparison of the data for the mapping points (Vhareta et al. (2014)) shows that for the earlier phase of limited fatigue cycles, there are relatively few points in which the local stress is tensile and that the number increases substantially when the samples have been subjected to a large number of cycles. These measurements confirm the differing nature of the mapping points and hence indicate that the fatigue cycles produce localized damage.

In particular, the results show the presence of an average compressive stress in all samples prior to the fatigue cycling which is initially reduced and ultimately leads to an average tensile stress when a sufficiently large number of fatigue cycles are applied.

**Fig. 13** Average surface stress variations as a function of the number of cycles for samples B, E, F, and G. Reproduced from Diam Relat Mater 45 (2014) 34–42, with the permission of Elsevier B.V. (<https://doi.org/10.1016/j.diamond.2014.02.012>)



**Fig. 14** The number of mapping points exhibiting tensile stress as a function of the number of fatigue cycles for samples B, E, F, and G. Reproduced from Diam Relat Mater 45 (2014) 34–42, with the permission of Elsevier B.V. (<https://doi.org/10.1016/j.diamond.2014.02.012>)



As indicated earlier, the diamond particles in PCD are bonded together by diamond-diamond growth taking place during the sintering process. The diamond particles thus form an intergrown matrix with pools and small channels of cobalt dispersed through this matrix. The diamond particles form the bulk of the PCD with the result that the PCD has hardness and wear resistance properties close to that of diamond. The Raman spectroscopy technique does not detect the stress state of the sintering aid and binder material, namely, the cobalt present in pores between the diamond grains. It is very likely that the cobalt is deformed plastically (i.e., beyond its yield limit) during the fatigue cycling at room temperature, thereby influencing the resultant stress state of the interlocking diamond matrix. The conditions for crack formation are enhanced by the presence of tensile stresses; thus, it is possible that progressive breaking of the inter-grain bridges associated with diamond to diamond bonding occurs during the fatigue cycles. The total resultant stress state of the PCD disk is therefore dependent on the distribution of component stress states between the diamond matrix and the cobalt binder phase.

The samples used in the present work display variations in their properties even though they were manufactured in the same batch. The average surface stresses of the individual samples show a considerable variation from one sample to the next. Some samples with a lower number of fatigue cycles exhibited an average tensile stress and a number of tensile mapping points greater than those with a higher number of cycles. These variations are considered to arise from a combination of sources. The compositional variations shown by the EDS results indicate differing conditions during their batch synthesis. An example would be temperature gradients at the time of sintering. Furthermore, the hardness of the diamond grains is highly dependent on grain size and orientation (Lammer (1988)). It is known from the use of PCD tool-bits that failure under practical conditions occurs erratically and the present results suggest that sample variability is the probable origin.

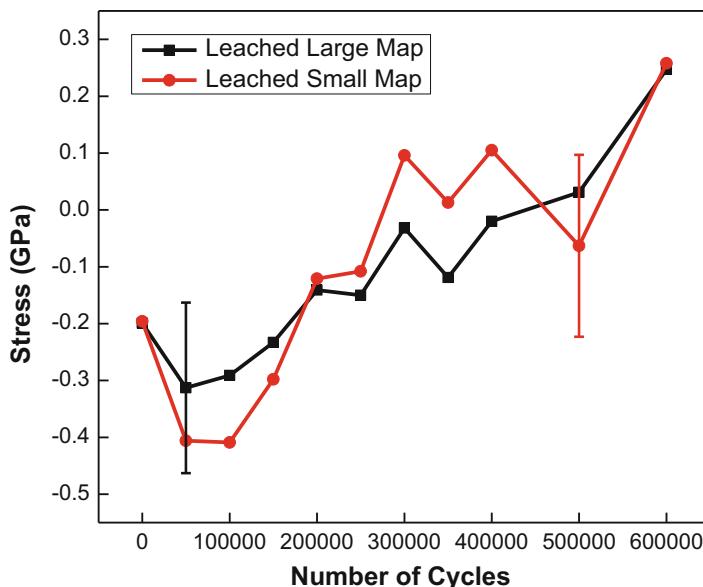
In an extension of the work discussed in the previous paragraphs, Raman spectroscopy and X-ray diffraction techniques were used on the same set of fatigued PCD samples. XRD is a widely used technique for nondestructive quantification of residual stress and strain in materials. A comparison of results of the two techniques was expected to yield valuable results.

The  $\sin^2\psi$  method was used for the XRD stress evaluation, where the lattice spacing at different tilt angles  $\psi$  is measured. The residual stress  $\sigma_\phi$  can be calculated from (see, e.g., Cullity and Weymouth 1978)

$$\sigma_\phi = \frac{E}{(1 + \nu) \sin^2\psi} \frac{d_\psi - d_0}{d_0}, \quad (38)$$

where  $d_\psi$  is the lattice spacing at a given  $\psi$  angle,  $d_0$  is the unstressed lattice spacing,  $E$  is Young's modulus, and  $\nu$  is Poisson's ratio. This equation allows for the calculation of the stress in any chosen direction from plane spacing and is valid whether the stress system is uniaxial or biaxial.

The fatigue method employed was the same as that described in Vhareta et al. (2014), except that the Co binder material was leached from the top 100 microns or so of the PCD disk. The results obtained via Raman spectroscopy was very similar to that reported previously. These results are given in Fig. 15 and show the gradual decrease of the magnitude of the average residual compressive stress and the eventual onset of average surface tensile stress with increasing number of fatigue

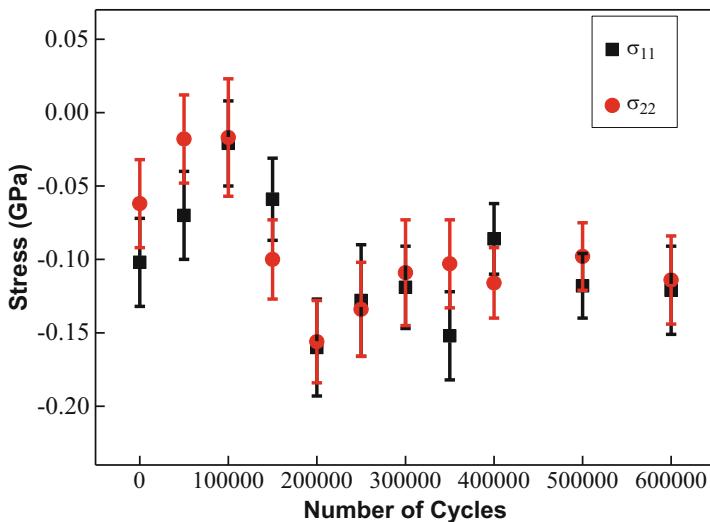


**Fig. 15** Average residual stress as measured on the surface of a PCD disk via Raman spectroscopy as a function of number of fatigue cycles

cycles. “Large map” and “small map” refer to the size of the mapping grid used relative to the size of the sample. “Large map” covers the largest possible surface area, whereas “small map” is confined to the central part of the disk (see Vhareta et al. (2014) for further details).

Figure 16 reports the residual stress as measured via the XRD technique. The two in-plane components of the stress are plotted separately for the same sample as measured in Fig. 15. As can be seen, the residual stress state at zero cycles is compressive in nature, although slightly smaller than the magnitude reported via Raman spectroscopy. There is essentially no change in the measured residual compressive stress state and magnitude with increasing number of fatigue cycles.

This was initially somewhat puzzling, but a literature survey showed that authors had reported similar results, albeit not on PCD. Ferreira et al. (2003) and Kim and Yu (1998) both noted differences in residual stresses measured via Raman spectroscopy and XRD in their work on diamond and diamond films. After some consideration, it was possible to explain the apparent differences in the results from the two methods by considering the penetration depth and the diameter of the sampling beam (spot size on the sample). The Raman technique makes use of visible radiation (typically in the green or blue part of the visible spectrum). The penetration depth of the laser beam into the PCD is thus limited by the binder matrix and grain boundaries to between 1 and 20  $\mu\text{m}$ . With Raman spectroscopy, only a fairly thin surface section of the material is thus probed. For the XRD technique, the penetration depth (sampling volume) is determined largely by the X-ray absorption coefficient of the sample material. For typical Cu K $\alpha$  radiation, the penetration depth for PCD is approximately 200  $\mu\text{m}$ . The X-ray technique thus samples a significantly “thicker” surface



**Fig. 16** Average residual stress as measured on the surface of a PCD disk via XRD as a function of number of fatigue cycles

layer. If the change in residual stress values is confined to a thin surface layer, as suggested by the Raman technique, then the XRD technique would not easily measure this, as the results would be dominated by the mostly constant compressive residual stress present in the thicker volume of the PCD.

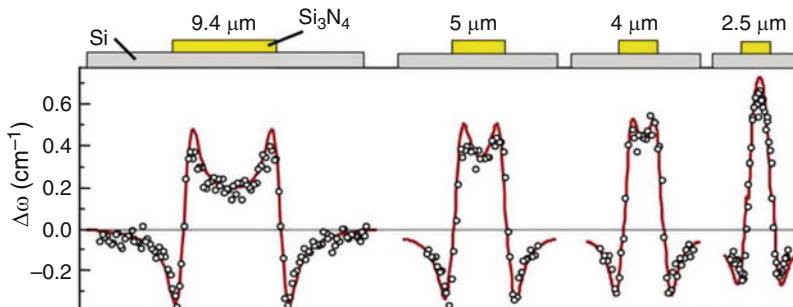
In terms of spot size, it is typically around 2  $\mu\text{m}$  in the Raman measurements reported here. It thus samples at the level of a single grain or intragranular. This is one of the reasons multiple points are measured and averaged. In the XRD technique, stress values obtained are the arithmetic average stresses in an area defined by dimensions of the beam size and diffraction angles range. In the XRD results reported here, the X-ray spot size was  $\sim 0.8 \text{ mm}$ . The X-ray beam is thus sampling multiple grains at a time, which arguably gives residual stress results over a different length scale compared to Raman spectroscopy (i.e., inter-granular as opposed to intra-granular).

Thus when taking these aspects into account, it is possible to reconcile the apparent differences in the results. It very much highlights the importance of taking the technique details and capabilities into account when evaluating important materials parameters such as stress and strain.

### Silicon Examples

Many examples of the application of Raman spectroscopy for the quantification of stress and strain in silicon and silicon-based devices are available. As an illustration, the work by De Wolf (1996, 2003, 2015) is briefly summarized here. The 1996 paper is a comprehensive topical review; the 2003 paper is a shorter overview; and the 2015 paper highlights the importance of the consideration of vertical stress components in modern 3D microelectronics. Different ways to calculate the relation between Raman frequency and triaxial stress, and the related Raman peak intensities, are discussed in detail in the latter paper.

Figure 17 shows how the position of the first-order Raman Si peak shifts as a function of stress due to the deposition of long  $\text{Si}_3\text{N}_4$  stripes with different widths onto an Si substrate. These nitride lines are under tensile stress due to the deposition



**Fig. 17** Shift of Raman silicon peak ( $\Delta\omega$ ) measured along nitride lines with different widths on a silicon substrate. The schematic rectangles indicate the position of the nitride lines and their widths (De Wolf 2003)

process. As a result, the Si atoms in the substrate under the stripes are compressed, causing compressive stress, and the Si atoms next to the lines are “pulled,” causing tensile stress. The process of the measurement and the fitting of an edge force model (solid lines in Fig. 17) to extract stress values from the measured Raman data is given in De Wolf (2003).

### Thin Films

The comments from section “[Collection Optics](#)” regarding penetration depth and sampling volume of the laser beam are particularly relevant to thin films. If the films are opaque to the incident radiation, generally only the Raman spectrum of the film is measured, but if the film is very thin, or has some transparency to the incident beam, the measured spectrum has peaks from the film and the substrate. With an appropriate choice of experimental conditions, a depth profile of, e.g., composition can be made for a stack of transparent film layers. A few review-type papers that address different aspects of the application of Raman spectroscopy to thin films are highlighted below.

The paper by Hou et al. (2011) discusses the use of Raman spectroscopy to study stresses that develop in thermally grown oxide layers, particularly  $\text{Cr}_2\text{O}_3$ , at higher temperatures (700 °C to 900 °C). It highlights several factors that, in addition to stress, affect the Raman peak shift position, such as temperature, non-stoichiometry, impurities, and instrumental calibration. It concludes with a discussion of the advantages and disadvantages of Raman spectroscopy for the study of stresses in oxide materials.

Studies of strain in single-walled carbon nanotubes (SWCNTs) were published by, e.g., Liu et al. (2009). The Raman spectrum of SWCNTs shows a rich diversity of peaks, related to the diameter and chirality of the nanotube. When external stress and strain is applied to a nanotube, its geometrical structures change, with consequent changes to its electronic structure and phonon properties. This paper studies the effect of the application of several distinct types of strain, including uniaxial strain, torsional strain, radial deformation, and bending deformation, on the Raman spectrum of SWCNTs.

Mohiuddin et al. (2009) studied the Raman spectrum of graphene as a function of uniaxial strain. They find that the doubly degenerate  $E_{2g}$  optical phonon mode splits into two components – one polarized along the strain and the other perpendicular to it. This is analogous with the effect of curvature on the G-peak of carbon nanotubes. These and other presented results show how Raman spectroscopy can be used to measure the magnitude of uniaxial or biaxial strain present, making this of particular interest in the field of carbon nanoelectronics.

Residual stresses and stress fields around cracks in undoped and phosphorus-doped homoepitaxial diamond films are studied using confocal micro-Raman scattering in the paper by Mermoux et al. (2005). Due to the confocal arrangement of the optics, the Raman signal arising from the grown film could be discriminated from that of the underlying diamond substrate. Films with (111) orientation exhibited substantial tensile stress, indicated by a downshift in the Raman peak and by spontaneous cracking in films thicker than 5 μm. A high compressive stress

was detected at the substrate near the surface, and a partial stress relaxation was observed in the vicinity of cracks. Gries et al. (2007) present calculations of phonon deformation potentials in textured cubic films for different directions of growth, applied to diamond. This is applied to studying stresses in polycrystalline diamond samples.

### Inclusions in Crystals

Where a host crystal is transparent to the laser wavelength being used for Raman spectroscopy, identification of inclusions and/or mapping of stress fields around an inclusion is possible. This usually requires a confocal Raman microscope to precisely locate the volume of material being sampled.

A very good example is given in the paper by Noguchi et al. (2013), where the internal stress field around a zircon inclusion in a sapphire crystal is mapped using both Raman and photoluminescence (PL) spectroscopy. It was found that the internal stress was highly concentrated at the tips of the zircon crystal, where the deviatoric stress and the hydrostatic pressure component reached 700 MPa and 470 MPa, respectively. The generation of compressive stress at the surface of the zircon crystal can be explained by the difference in thermal expansion coefficients and elastic constants between sapphire and zircon. In general, internal stress fields induced by mineral inclusions reflect the pressure and temperature conditions at which the host sapphire gemstones were crystallized and the paper thus illustrates the use of Raman and PL spectroscopy to identify the origin of gemstones.

In a related paper, Korsakov et al. (2015) used confocal 2D and 3D Raman imaging of metamorphic diamonds to reveal their complex internal morphology. The growth patterns of quite small diamonds, less than 10 µm in size, can be measured, and this in conjunction with techniques such as cathodoluminescence reveals the distribution of defects giving rise to the observed luminescence.

It is also mentioned here that Raman spectroscopy has wide application in the study of fluid inclusions in transparent mineral hosts. Its use ranges from qualitative detection of solid, liquid, and gaseous components to identification of polyatomic ions in solution. From the Raman data, the density of CO<sub>2</sub> fluids, the chemistry of aqueous fluids, and the molar proportions of gaseous mixtures present in the inclusions can be determined. It can also be used to measure the pH range and oxidation state of fluids. In these applications, the main advantages of the technique are minimal sample preparation, high versatility, and nondestructive evaluation of the samples. Reviews in this regard can be found by Burke (2001) and Frezzotti et al. (2012).

### Applications Using Phase Identification

One of the most frequent applications of Raman spectroscopy is in the identification of phases of materials and also the tracking of phase changes as, e.g., functions of temperature or pressure (or both). An often used example is that of carbon: in its 3D bonded morphology, it is diamond, while in its 2D morphology, it is graphite. The two

materials have very distinct Raman spectra due to the differences in bonding lengths and bonding angles of the C-C bonds in the two materials. Phase identification thus finds wide application in physics, chemistry, geology, and many other associated disciplines.

### Raman Spectroscopy in Corrosion Processes

Here the phase identification application of Raman spectroscopy is illustrated in the study of the growth and localized breakdown of the passive film on iron during corrosion processes. The oxidation of iron atoms resulting in corrosion of a metal surface is thermodynamically favored and driven by reduction of oxygen and water in air. Addition of certain metals (e.g., Ni, Cr, Mo) forms a thin oxide film on the metal surface, thus retarding the rate of further oxidation and passivating the surface to form stainless steels. However, passivated metals tend to undergo pitting; this, being localized corrosion, results in construction and mechanical failures (Stansbury and Buchanan 2000; Sato 1997; Cohen 1978). Nieuwoudt et al. (2011a, b) developed improved *in situ* Raman spectroscopic methods in the study of these processes including the composition of the passive film formed on pure iron and the changes of the film during pitting under the influence of chlorine ions.

In the Raman studies, a Spectra-Physics 365/35 tunable dye laser used with R6G dye and pumped by a Spectra-Physics 171 argon ion laser was used to determine an optimized excitation wavelength. A study of the composition of the passive film on iron in a 0.05 M NaOH solution increased the intensity of the Raman bands of the iron oxide and oxyhydroxide compounds by as much as a factor of ten with the laser excitation at 636.4 nm rather than the conventional 514.5 nm line of an argon ion laser. This is due to pre-resonance enhancement in which the excitation wavelength is close to the 640 nm absorption band resulting from ligand field transitions of octahedrally coordinated Fe<sup>3+</sup>.

The complex, multicomponent spectra obtained *in situ* from the various iron oxide and oxyhydroxide compounds of the respective passive films were resolved into their individual components using multivariate curve resolution with alternating least squares (MCR-ALS) (Zakroczymski et al. 1985). Thus the composition of the passive films could be determined during their growth or localized breakdown.

Since water plays a prominent role in the passive film, similar methods were used to determine the behavior of water within the surface region using an excitation wavelength of 514.5 nm. Characterization of the composition of the iron oxide compounds and monitoring of the changes of the water present in the film in the 0.05 M NaOH solution at different stages of pitting under the influence of chlorine ions provided an understanding of the mechanism of pitting of iron (Jaumot et al. 2005; Sato 1989).

The controlled growth of the passive film on iron was achieved by potentiodynamic polarization employing potentiostats. (Bioanalytical Systems CV27 for passive film growth and Schlumberger SII286 for pitting investigations.)

A three-electrode electrochemical cell was constructed to be placed on the stage of an Olympus confocal Raman microscope attached to a Jobin-Yvon T64000 Raman spectrograph. The surface of the iron sample within the cell was analyzed using a backscattering geometry with a long working distance 20 X microscope

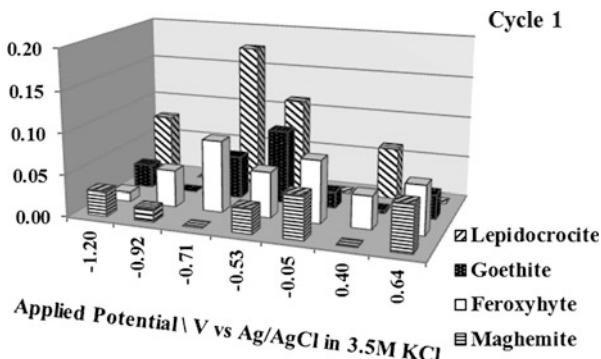
objective. More detailed descriptions of the electrochemical conditions and the three-electrode cell design are given in Nieuwoudt et al. (2011a, b). The Raman spectrograph was used in the single spectrograph mode with holographic notch filters for the rejection of the Rayleigh scattered light.

With the three-electrode cell, the temperature and flow rate of the electrolyte could be varied and Raman spectra were recorded *in situ* at selected values of the applied potential while the potential at the iron surface was controlled by the potentiostat. This determined the composition of the passive film at the different stages during the polarization sweep.

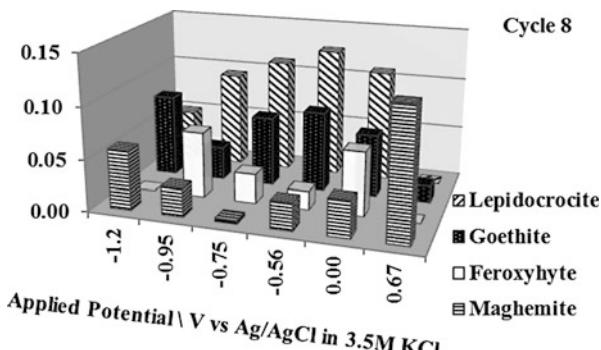
For the study of the composition of the passive film during its growth by successive potentiodynamic polarization cycles in 0.05 M NaOH, 18 cycles of anodic and cathodic sweeps were applied to the iron surface at 10 mV/s between -1.20 V and +0.67 V. During the 1st, 2nd, 5th, 6th, 8th, and 15th cycles, the scan was interrupted at various selected potentials while Raman spectra of the iron surface were recorded. MCR-ALS analyses of the Raman spectra carried out during cycles 1, 8, and 15, which represent the early and latter stages of growth, are shown below (Figs. 18, 19, and 20).

In order to monitor the composition of the passive film after addition of chlorine ions to the 0.05 M NaOH solution, the iron surface was prepared and mounted as the working

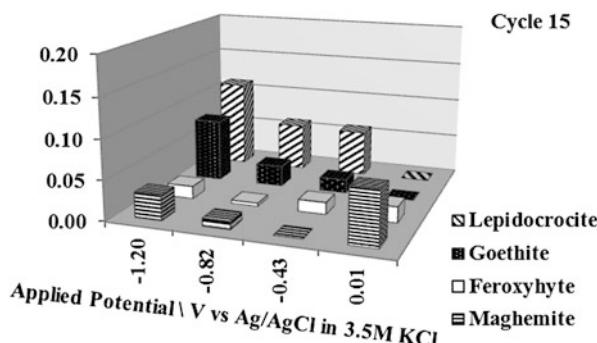
**Fig. 18** Relative amounts of iron oxide compounds determined by MCR-ALS analysis of the Raman spectra recorded during the anodic sweep of cycle 1. Reproduced from J. Raman Spectrosc. 42 (2011) 1353–1365, with the permission of John Wiley and Sons (<https://doi.org/10.1002/jrs.2842>)



**Fig. 19** Relative amounts of iron oxide compounds determined by MCR-ALS analysis of the Raman spectra recorded during the anodic sweep of cycle 8. Reproduced from J. Raman Spectrosc. 42 (2011) 1353–1365, with the permission of John Wiley and Sons (<https://doi.org/10.1002/jrs.2842>)



**Fig. 20** Relative amounts of iron oxide compounds determined by MCR-ALS analysis of the Raman spectra recorded during the anodic sweep of cycle 15. Reproduced from J. Raman Spectrosc. 42 (2011) 1353–1365, with the permission of John Wiley and Sons (<https://doi.org/10.1002/jrs.2842>)

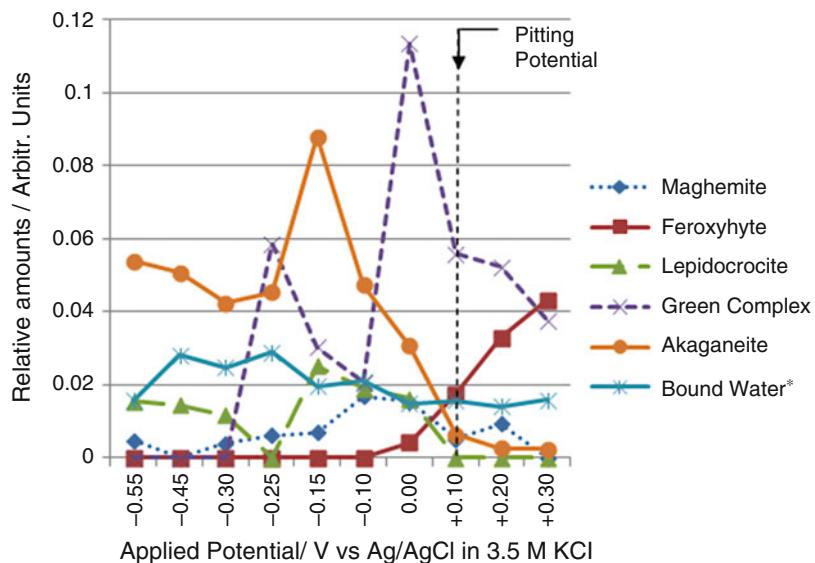


electrode in the three-electrode cell and connected to the Schlumberger SI1286 potentiostat. The surface was then subjected to cathodic polarization followed at  $-1.40$  V for 20 min. in 0.05 M NaOH. This was followed by two full polarization cycles between  $-1.20$  V and  $+0.60$  V at a scan rate of 10 mV/s in order to grow the passive film on the surface. During the third anodic sweep, the potential was held at  $-0.55$  V being the beginning of the passive region. At this point, a solution of NaCl was added to the electrolyte such that its concentration was 0.05 M so that the resulting electrolyte solution was 0.05 M NaOH + 0.05 M NaCl with pH 12.7. The potential was then increased in increments of 0.1 V up to  $+0.3$  V, and held there for 1200 s after each increment while the current was measured and Raman spectra of the surface were recorded.

Pitting events were manifested by short, sharp current spikes in chronoamperometric measurements. Sudden and constant increases in the current indicated the onset of stable pitting occurring after a 1000 s induction time while the potential was held at  $+0.10$  V. From the MCR-ALS analysis of the Raman spectra recorded in situ at each potential increment between  $-0.55$  V and  $+0.3$  V, and also of the pits and their adjacent surfaces at the pitting potential, the compounds present in the passive film were determined to be maghemite  $\gamma\text{-Fe}_2\text{O}_3$ , feroxyhyte  $\delta\text{-FeOOH}$ , lepidocrocite  $\gamma\text{-FeOOH}$ , akaganeite  $\beta\text{-FeOOH}$ , and green complex, the latter being a more hydrated amorphous magnetite compound than green rust. The results are shown in Fig. 21.

Raman spectra of the passive film in the O-H stretch region were recorded in situ using the argon ion laser with the wavelength of 514.5 nm. The Raman spectra recorded the O-H stretch vibration between 4300 and 2300  $\text{cm}^{-1}$  at the same potentials as used previously and were used to determine the amount of water bound in the passive film as the potential was increased up to and beyond the pitting potential. Figure 21 shows that more water becomes incorporated into that film as the surface begins to passivate at  $-0.55$  V and then begins to decrease further into the passive region, at around 0.0 V.

The composition of each of the pits which had been formed at  $+0.1$  and  $+0.3$  V was mainly magnetite ( $\text{Fe}_3\text{O}_4$ ) and feroxyhyte ( $\delta\text{-FeOOH}$ ). The composition of the pit components and their adjacent surfaces at the pitting potentials were also determined from the in situ spectra recorded at  $+0.1$  V and are shown in Fig. 22.

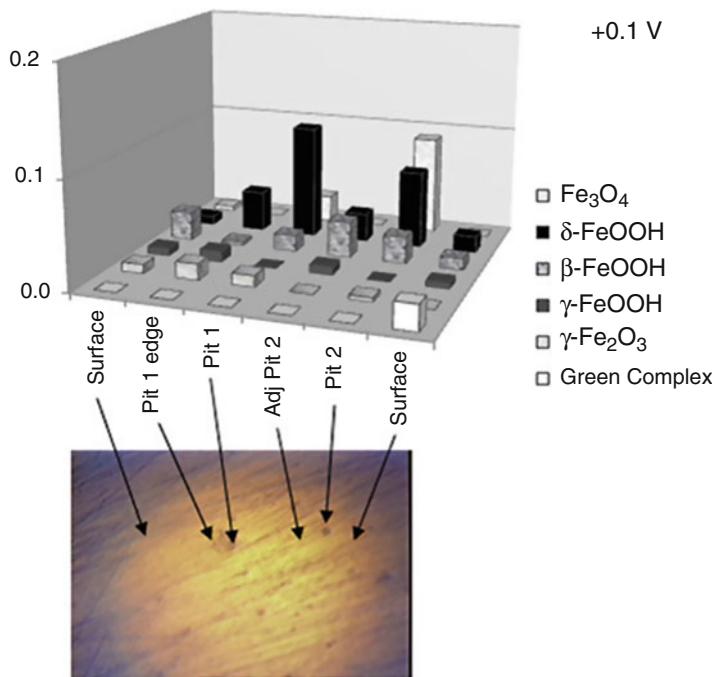


**Fig. 21** Relative amounts of maghemite, feroxyhyte, lepidocrocite, akaganeite, green complex, and bound water in the passive film determined using MCR-ALG optimization at different potentials after addition of  $\text{Cl}^-$  to the 0.05 M NaOH solution at  $-0.55$  V. Bound water amounts were scaled down by a factor of 10 for clearer comparison of the trend

The composition of the pits was mainly feroxyhyte and magnetite with relatively smaller amount of maghemite ( $\gamma\text{-Fe}_2\text{O}_3$ ) and akaganeite ( $\beta\text{-FeOOH}$ ). The surfaces with no pits and adjacent to the pits contained more or less equal amounts of green complex, feroxyhyte, akaganeite, maghemite, and lepidocrocite ( $\gamma\text{-FeOOH}$ ). Lepidocrocite or green complex was found in the pits, and no magnetite was present in the surfaces adjacent to the pits.

Around the pitting potential, a decrease in the amount of water incorporated in the surface occurred along with a change in the composition of the passive film: specifically a decrease in some of the forms of iron oxides with formation of another. It is considered that water plays a protective role in the passive film.

In summary, using pre-resonance enhancement of the Raman spectra and MCR-ALS, it was possible to identify the different FeOOH compounds on the surface and to distinguish more clearly between  $\text{Fe}_3\text{O}_4$  and  $\gamma\text{-Fe}_2\text{O}_3$  during the initial stages of the passive film on iron by cyclic polarization. The results corroborate the two-layer model for the passive film on iron, with  $\gamma\text{-Fe}_2\text{O}_3$  forming the inner layer and the outer layer consisting of  $\delta\text{-FeOOH}$ ,  $\alpha\text{-FeOOH}$ ,  $\gamma\text{-FeOOH}$ , and other components such as  $\text{Fe(OH)}_2$  and other intermediates. The amounts of these components vary with potential and cycling. In the passive region of anodic polarization, the hydration of the films become enhanced with increasing anodic potential and with a larger number of cycles becomes amorphous, comparing well with the hydrated polymeric oxide model for the passive film.



**Fig. 22** Relative amounts of iron oxide components and green complex determined using MCR-ALS optimization from spectra recorded *in situ* from pits and the surrounding areas on the iron surface during stable pitting at +0.1 V. Reproduced from J. Raman Spectrosc. 43 (2012) 928–938, with the permission of John Wiley and Sons (<https://doi.org/10.1002/jrs.3109>)

### Other Examples

As a further illustration, the paper by Sherif et al. (2010) studies corrosion and corrosion inhibition of single crystal pure iron in a 3.5% sodium chloride solution using Raman spectroscopy and electrochemical techniques. The corrosion inhibitor used was 3-amino-5-mercaptop-1,2,4-triazole (AMTA). A range of electrochemical techniques showed that the presence of AMTA and an increase of its concentration moves the corrosion potential to more positive values and decreases both the corrosion current and corrosion rate. *In situ* and *ex situ* Raman investigations confirmed that the addition of AMTA molecules to the chloride solution strongly inhibits the iron corrosion through their adsorption onto the surface, blocking its active sites and preventing its corrosion.

As mentioned earlier, luminescence spectroscopy can be used in combination with Raman spectroscopy to increase the information that can be gathered from a sample. As an example of this, the paper by Odusote et al. (2013) measures residual stress in alumina scale by photostimulated  $\text{Cr}^{3+}$  luminescence piezo-spectroscopy. The shift of the position of the R-line luminescence doublet from  $\text{Cr}^{3+}$  is used to calculate the residual stress, assuming no particular crystallographic texture and a state of biaxial compression.

## Comments on Other NDE Applications and Technique Developments

### Other NDE Applications

As stated earlier, the applications and examples shown in this chapter have focused mainly on inorganic and engineering type materials. A few papers to give an indication of the scope of other NDE-related work are mentioned here.

Dooley et al. (2009) report on “stress mapping of undamaged, strained, and failed regions of bone using Raman spectroscopy.” Pezzotti (2007) reviews stress microscopy and confocal Raman imaging of load-bearing surfaces in artificial hip joints. Raman spectroscopy in cancer studies is widely reported; see, for example, the review paper by Austin et al. (2016), where Raman technologies in cancer diagnostics are discussed. Spontaneous, coherent and surface-enhanced Raman spectroscopies and imaging, as well as the fundamental principles governing the successful use of these techniques, are covered and the utilization of the discussed Raman techniques for the detection and diagnosis of cancer *in vitro*, *ex vivo*, and *in vivo* is described. The review concludes with a discussion of the future directions of Raman technologies, with particular emphasis on their clinical translation. Ishigaki et al. (2017) discuss the nondestructive monitoring of mouse embryo development and its qualitative evaluation at the molecular level using Raman spectroscopy.

Raman spectroscopy has had wide application as an NDE technique in art and archaeology, with whole books having been published on the subject. For recent reviews, see e.g., Edwards and Vandenabeele (2016) and Colomban (2017). Both papers are of an overview nature, written by experts in the fields, and thus contain many useful references therein.

Papers by Wei et al. (2015) and Foucher et al. (2015) both highlight the potential of Raman spectroscopy in extra-terrestrial research. The first paper discusses testing of a Raman instrument intended to be installed on a future Mars rover, and the second discusses Raman mapping as a potential method for the detection of micro-fossils on Mars.

### Technique Developments

A few comments are made here about specific Raman spectroscopy technique developments. The technique of surface-enhanced Raman spectroscopy (SERS) is a phenomenon in which the Raman signals of molecules are enormously enhanced when they are very close to certain SERS-active nanostructures. The paper by Panneerselvam et al. (2018) (and references therein) contains a detailed discussion about the mechanisms, applications, and current challenges. In summary, nanostructures of free-electron-like metals (e.g., Au, Ag, and Cu) are able to massively concentrate incident optical fields at locations where the analyte of interest is present and could also enhance the radiation efficiency of the oscillating dipole source of the analyte at the Raman scattering frequencies. This two-step enhancement is mainly due to the excitation of surface plasmons (SPs) of the metal nanostructures by incident light in the far field and Raman scattering of oscillating dipole radiation

in the near field. Through this two-step enhancement process, the total Raman intensities of analytes can usually be increased by up to 6 to 8 orders of magnitude. The discovery of SERS soon created an upsurge of interest, as SERS can detect ultralow concentrations of target molecules on plasmonic metal nanostructures, even at a single-molecule level. Thus, SERS has attracted widespread interest owing to its ultrahigh sensitivity and molecular specificity in chemical analysis, surface analysis, and biomolecular analysis.

Tip-enhanced Raman spectroscopy (TERS) provides an interesting alternative for going beyond the diffraction limit by performing localized Raman spectroscopic scanning or mapping with high spatial resolution on a wide range of nanomaterials such as carbon nanotubes and graphene, self-assembled organic monolayers, and DNA molecules. This is possible because of the local amplification and the spatial confinement of the electromagnetic radiation by the apex of a sharp metallic tip, usually that of an atomic force microscope (AFM). TERS can thus offer unprecedented detail in the spatially resolved phase composition and strain mapping of single nanostructures, without requiring destructive procedures for sample preparation. The paper by Tarun et al. (2009) provides a review of nanoscale strain characterization with a focus on strain in silicon nanodevices. The paper by Yeo et al. (2009) offers an overview discussion of TERS with a focus on tip plasmonics, single molecule detection, and nanoscale chemical analysis of biological samples. As a more specific example of highly localized strain in nanostructures, the letter by Rahaman et al. (2017) reports on TERS imaging of ultrathin MoS<sub>2</sub> flakes giving strain evidence of strain changes spatially localized to regions as small as 25 nm.

---

## Summary

The results presented here in conjunction with the discussions illustrate clearly of what benefit the optical spectroscopy technique of Raman scattering is to the NDE of hard materials. Raman spectroscopy has the ability to clearly map stress and deformation distributions in materials, as illustrated for single crystal diamond and PCD. Application to corrosion processes on the surface of iron shows how the evolution of iron oxide and -oxyhydroxide products can be monitored and conclusions about mechanism can be made. The relatively good spatial resolution inherent in the technique makes it possible to study corrosion mechanisms at individual pits. It is anticipated that Raman scattering techniques will continue to play an ever-increasing role in research in a wide range of disciplines, which includes NDE in multiple fields.

**Acknowledgements** Financial support from the National Research Foundation (NRF) of South Africa under Grant No 2053306, the DST-NRF Centre of Excellence in Strong Materials hosted by the University of the Witwatersrand, the University of the Witwatersrand, Johannesburg, and the African Laser Centre (ALC) is gratefully acknowledged.

## References

- Adar F, Mamedov S, Whitley A (2010) *Microsc Microanal* 16:360–361
- Ager JW, Drory D (1993) *Phys Rev B* 48:2601–2607
- Ager JW, Veirs DK, Rosenblatt GM (1991) *Phys Rev B* 43:6491–6499
- Anastassakis E, Pinczuk A, Burstein E, Pollak FH, Cardona M (1970) *Solid State Commun* 8:133–138
- Austin LA, Osseiran S, Evans CL (2016) *Analyst* 141:476–503
- Balkanski M, Wallis RF, Haro E (1983) *Phys Rev B* 28:1928–1934
- Beechem T, Graham S, Kearney SP, Phinney LM, Serrano JR (2007) *Rev Sci Instrum* 78:061301
- Bergman L, Nemanich RJ (1995) *J Appl Phys* 78:6709–6719
- Boppert H, van Straaten J, Silvera IF (1985) *Phys Rev B Rapid Commun* 32:1423–1425
- Borer WJ, Mitra SS, Namjoshi DV (1971) *Solid State Commun* 9:1377–1381
- Boteler JM, Gupta YM (1993) *Phys Rev Lett* 71:3497–3500
- Brookes CA (1992) In: Field JE (ed) *The properties of natural and synthetic diamond*. Academic, London, p 515
- Brookes CA, Brookes EJ, Howes VR, Roberts SG, Waddington CP (1990) *J Hard Mater* 1:3–24
- Brüesch P (1982) *Phonons: theory and experiments I*. Springer, New York
- Brüesch P (1986) *Phonons: theory and experiments II*. Springer, Berlin Heidelberg
- Burke EAJ (2001) *Lithos* 55:139–158
- Campbell IH, Fauchet PM (1986) *Solid State Commun* 58:739–741
- Cardona M (1982) In: Cardona M, Güntherodt G (eds) *Light scattering in solids II (topics in applied physics 50)*. Springer, Berlin
- Catledge SA, Vohra YK (1995) *J Appl Phys* 78:7053–7058
- Catledge SA, Vohra YK, Ladi R, Rai G (1996) *Diam Relat Mater* 5:1159–1165
- Cohen M (1978) In: Frakenthal RP, Kruger J (eds) *Passivity of metals*. The Electrochemical Society Inc, New Jersey, pp 521–545
- Colomban P (2002) *Adv Eng Mater* 4:535–542
- Colomban P (2017) *J Raman Spectrosc* 2017:1–14
- Colomban P, Gouadec G, Mathez J, Tschiember J, Pérès P (2006) *Compos Part A-Appl S* 37:646–651
- Crawford FS Jr (1968) *Waves: Berkeley physics course, vol 3*. McGraw-Hill, New York, p 376
- Cullity BD, Weymouth JW (1978) *Elements of X-ray diffraction*, 2nd edn. Addison-Wesley, New York
- Datchi F, Canny B (2004) *Phys Rev B* 69:144106
- de la Vega A, Kinloch IA, Young RJ, Bauhofer W, Schulte K (2011) *Compos Sci Technol* 71:160–166
- De Wolf I (1996) *Semicond Sci Tech* 11:139–154
- De Wolf I (2003) *Spectrosc Eur* 15/2:6–13
- De Wolf I (2015) *J Appl Phys* 118:053101
- De Wolf I, Anastassakis E (1999) *J Appl Phys* 85:7484–7485
- De Wolf I, Maes HE, Jones SK (1996) *J Appl Phys* 79:7148–7156
- Dooley KA, McCormack J, Fyhrie DP, Morris MD (2009) *J Biomed Opt* 14:044018
- Edwards HGM, Vandenebeele P (2016) *Philos T Roy Soc A* 374:20160052
- Erasmus RM, Comins JD, Fish ML (2000) *Diam Relat Mater* 9:600–604
- Erasmus RM, Daniel RD, Comins JD (2011a) *J Appl Phys* 109:013527
- Erasmus RM, Comins JD, Mofokeng V, Martin Z (2011b) *Diam Relat Mater* 20:907–911
- Evans T, Davey ST, Robertson SH (1984) *J Mater Sci* 19:2405–2414
- Everall N (2010) *Analyst* 135:2512–2522
- Everall N (2014) *J Raman Spectrosc* 45:133–138
- Ferreira NG, Abramof E, Corat EJ, Trava-Airoldi VJ (2003) *Carbon* 41:1301–1308
- Field JE (1992) In: Field JE (ed) *The properties of natural and synthetic diamond*. Academic Press, London, p 667

- Foucher F, Ammar M-R, Westall F (2015) *J Raman Spectrosc* 46:873–879
- Frezzotti ML, Tecce F, Casagli A (2012) *J Geochem Explor* 112:1–20
- Ganesan S, Maradudin AA, Oitmaja J (1970) *Ann Phys-New York* 56:556–594
- Gries T, Vandenbulcke L, Simon P, Canizares A (2007) *J Appl Phys* 102:083519
- Griffith WP (1975) In: Karr C (ed) *Infrared and Raman spectroscopy of lunar and terrestrial minerals*. Academic Press, New York
- Grimsditch MH, Anastassakis E, Cardona M (1978) *Phys Rev B* 18:901–904
- Gupta YM, Horn PD, Yoo CS (1989) *Appl Phys Lett* 55:33–35
- Hanfland M, Syassen K, Fahy S, Louie SG, Cohen ML (1985) *Phys Rev B* 31:6896–6899 *Rapid Comm*
- Harris TK, Brookes EJ, Daniel RD (2001) *Diam Relat Mater* 10:755–759
- Hart TR, Aggarwal RL, Lax B (1970) *Phys Rev B* 1:638–642
- Herchen H, Cappelli MA (1993) *Phys Rev B* 47:14193–14199
- Hou PY, Ager J, Mougin J, Galerie A (2011) *Oxid Met* 75:229–245
- Imanaka M, Ishikawa R, Sakurai Y, Ochi K (2009) *J Mater Sci* 44:976–984
- Ishigaki M, Hashimoto K, Sato H, Ozaki Y (2017) *Sci Rep-UK* 7:43942
- Jaumot J, Gargallo R, de Juan A, Tauler R (2005) *Chemometr Intell Lab* 76:101–110
- Jothilakshmi R, Ramakrishnan V, Kumar J, Sarua A, Kuball M (2011) *J Raman Spectrosc* 42:422–428
- Kereszty G (2002) In: Chalmers JM, Griffiths PR (eds) *Handbook of vibrational spectroscopy, Theory and instrumentation*, vol 1. Wiley, Chichester
- Kim JG, Yu J (1998) *J Mater Res* 13:3027–3033
- Klein MV (1990) In: Horton GK, Maradudin AA (eds) *Dynamical properties of solids*. North-Holland, Amsterdam
- Korsakov AV, Toporski J, Dieing T, Yang J, Zelenovskiy PS (2015) *J Raman Spectrosc* 46:880–888
- Lammer A (1988) *Mater Sci Tech-Lond* 4:949–955
- Landsberg G, Mandelstam L (1928a) *Naturwissenschaften* 16:557–558
- Landsberg G, Mandelstam L (1928b) *Z Phys* 50:769–780
- Lee CJ, Pezzotti G, Okui Y, Nishino S (2004) *Appl Surf Sci* 228:10–16
- Liu Z, Zhang J, Gao B (2009) *Chem Commun* 2009:6902–6918
- Long DA (2002) *The Raman effect: a unified treatment of the theory of Raman scattering by molecules*. Wiley, New York
- Loudon R (1964) *Adv Phys* 13:423–482
- Marcuse D (1980) *Principles of quantum electronics*. Academic, New York
- McNamara D, Alveen P, Damm S, Carolan D, Rice JH, Murphy N, Ivanković A (2015) *Int J Refract Met H* 52:114–122
- Mermoux M, Marcus B, Crisci A, Tajani A, Gheeraert E, Bustarret E (2005) *J Appl Phys* 97:043530
- Mitra SS, Brafman O, Daniels WB, Crawford RK (1969) *Phys Rev* 186:942–944
- Mohiuddin TMG, Lombardo A, Nair RR, Bonetti A, Savini G, Jalil R, Bonini N, Basko DM, Gallois C (2009) *Phys Rev B* 79:205433
- Mohrbacher H, Van Acker K, Blanpain B, Van Houtte P, Celis J-P (1996) *J Mater Res* 11:1776–1782
- Mossbrucker J, Grotjohn TA (1997) *J Vac Sci Technol A* 15:1206–1210
- Muraki N, Katagiri G, Sergio V, Pezzotti G, Nishida T (1997) *J Mater Sci* 32:5419–5423
- Nafie LA (2017) *J Raman Spectrosc* 48:1692–1717
- Nemanich RJ, Solin SA, Martin RM (1981) *Phys Rev B* 23:6348–6356
- Nieuwoudt MK, Comins JD, Cukrowski I (2011a) *J Raman Spectrosc* 42:1335–1339
- Nieuwoudt MK, Comins JD, Cukrowski I (2011b) *J Raman Spectrosc* 42:1353–1365
- Noguchi N, Abduriyim A, Shimizu I, Kamegata N, Odakea S, Kagia H (2013) *J Raman Spectrosc* 44:147–154
- Nugent KW, Prawer S (1998) *Diam Relat Mater* 7:215–221
- Odusote JK, Cornish LA, Chown LH, Erasmus RM (2013) *Corros Sci* 70:276–284

- Panneerselvam R, Liu G-K, Wang Y-H, Liu J-Y, Ding S-Y, Li J-F, Wu D-Y, Tian Z-Q (2018) *Chem Commun* 54:10–25
- Parsons BJ (1977) *Proc R Soc Lon Ser-A* 352:397–417
- Pezzotti G (2007) *Expert Rev Med Devic* 4:165–189
- Postmus C, Ferraro JR, Mitra SS (1968) *Phys Rev* 174:983–987
- Rahaman M, Rodriguez RD, Plechinger G, Moras S, Schüller C, Korn T, Zahn DRT (2017) *Nano Lett* 17:6027–6033
- Raman CV (1928) *Indian J Phys* 2:387–398
- Raman CV, Krishnan KS (1928) *Nature* 121:501–502
- Richter H, Wang ZP, Ley L (1981) *Solid State Commun* 39:625–629
- Roberts SG (1988) *Philos Mag A* 58:347–364
- Sato N (1989) *Corros Sci* 31:1–19
- Sato N (1997) *Corros Sci* 27:421–433
- Sharma SK, Mao HK, Bell PM, Xu JA (1985) *J Raman Spectrosc* 16:350–352
- Sherif E-SM, Erasmus RM, Comins JD (2010) *Electrochim Acta* 55:3657–3663
- Srikar VT, Swan AK, Ünlü MS, Goldberg BB, Spearing SM (2003) *J Microelectromech S* 12:779–787
- Stansbury EE, Buchanan RA (2000) Fundamentals of electrochemical corrosion, 1st edn. ASM International (Publishers), Ohio
- Starman LA Jr, Lott JA, Amer MS, Cowan WD, Busbee JD (2003) *Sensors Actuat A-Phys* 104:107–116
- Tarun A, Hayazawa N, Kawata S (2009) *Anal Bioanal Chem* 394:1775–1785
- Thorne AP (1988) Spectrophysics, 2nd edn. Chapman and Hall, London
- Uehara K, Yamaya S (1988) *Int J Refract Met H* 7:219–223
- Van Camp PE, Van Doren VE, Devreese JT (1992) *Solid State Commun* 84:731–733
- Vhareta M, Erasmus RM, Comins JD (2014) *Diam Relat Mater* 45:34–42
- Wei J, Wang A, Lambert JL, Wettergreen D, Cabrol N, Warren-Rhodes K, Zacny K (2015) *J Raman Spectrosc* 46:810–821
- Whalley E, Lavergne A, Wong PTT (1976) *Rev Sci Instrum* 47:845–848
- Yeo B-S, Stadler J, Schmid T, Zenobi R, Zhang W (2009) *Chem Phys Lett* 472:1–13
- Zakroczymski T, Fan C-J, Szklarska-Smialowska Z (1985) *J Electrochem Soc* 132:2868–2871
- Zhao Q, Wagner HD (2004) *Philos T Roy Soc A* 362:2407–2424



# Optical Fiber Methods in Nondestructive Evaluation

18

Wolfgang R. Habel

## Contents

Introduction to Optical Fiber Sensor Technology .....	596
Optical Fiber Sensor Classes and Sensing Technologies .....	597
Sensors for Local Measurements .....	601
Multiplexed Single-Point and Segmented Sensors .....	611
Sensors for Distributed Measurements .....	612
Application Examples .....	615
Sensors for Local and Quasi-distributed Measurements .....	615
Sensors for Fully Distributed Measurements .....	631
Standardization .....	635
Summary and Future Trends .....	638
References .....	640

## Abstract

Optical fiber sensors are becoming increasingly significant as a smart sensing technology for NDE purposes and for monitoring of structures because of their extremely small dimensions of the sensing element and the leading optical fibers, their excellent static and dynamic measuring resolution, their resistance to many chemicals, and their immunity against high electromagnetic fields. This chapter introduces the basic features and physical principles of different types of optical fiber sensors used to evaluate the structure's behavior. The description covers all relevant sensor types mainly for measurement of mechanical quantities, but also for evaluation of functionality, based on the appropriate chemical and physical properties. The sections focus on optical fiber sensors applied to surfaces and embedded into materials to measure local strain changes and to evaluate

---

W. R. Habel (✉)

Department of Non-Destructive Testing, Division of Fibre Optic Sensors, BAM Federal Institute for Materials Research and Testing (formerly), Berlin, Germany  
e-mail: [wolfgang.habel@gmail.com](mailto:wolfgang.habel@gmail.com); [w.habel@ht.tu-berlin.de](mailto:w.habel@ht.tu-berlin.de)

acoustic emissions. Other types focus on the measurement of strain profiles over extended areas of structures or in hidden loaded structural components like anchors. Finally, distributed optical fiber sensors for the measurement of strain, temperature, and acoustic emissions at any (unknown) location along the length of long-gage optical fiber sensors are described. Typical examples from different fields of NDE and materials characterization including the capability in these fields to reveal severe damaging processes underline the usefulness of optical fiber sensors and illustrate the unique possibilities to achieve information about the behavior of structures and materials. The chapter is completed with a report on current international activities in standardization of optical fiber sensors and, finally, with a summary and some thoughts about future trends.

---

## Introduction to Optical Fiber Sensor Technology

Optical fibers are widely known as transport media for transmission of data and in telecommunication, where pulses have to be transmitted over long distances with high transmission rates. However, optical fibers can also be used as sensors. Glass and polymer optical fibers are sensitive to external mechanical and physical perturbations. In contrast to high-speed data communication, sensor technology experts exploit this sensitivity to develop sensors.

Optical fiber sensors (OFS) became increasingly significant as a smart sensing technology in the past two decades. The reasons are:

- Their capability of being very sensitive, small, lightweight and chemically inert, and without perturbing structural properties when embedded
- Their capability of being highly distributed
- Enable high static and dynamic resolution
- Withstand a few hundred degrees Celsius during the curing process of composites
- Do not need electrical power at the place of measurement
- Are electrically passive and not disturbed by electromagnetic fields or by parasitic currents
- Enable galvanic isolation, are intrinsically safe, and there is a low risk of sparking because of the very low radiant energy emerging from the optical fiber sensor system
- Cannot be attacked by lightning strikes
- Can be divided into several measuring sections, are network-compatible and amenable to multiplexing
- Have small interface requirements (the optoelectronic elements and demodulation electronics are confined in the reading unit)
- Are almost exclusively driven by standard photonics components

The OFS system consists of two parts:

- (a) The sensing unit: it contains the fiber-optic sensing element equipped with a protective coating. The sensing element can get additional layers or design

- elements to apply it to the measuring object; it can also be integrated in a small tube.
- (b) The optoelectronic unit: it contains the radiation source (semiconductor laser diode LD or a luminescence diode LED) and a photodetector (PIN diode or avalanche photodiode APD).

For NDE purposes, the greatest advantages of optical fiber sensors are the extremely small dimensions of the sensing element and the leading optical fibers, which avoid any perturbation in the measuring area and any influence to the integrity of the material, if the sensors are embedded. Another significant advantage of glass optical sensors is that they are resistant to almost all chemicals even at elevated temperatures. If NDE has to be performed under full environmental conditions, glass or polymer optical fibers are resistant to weathering effects and are not subject to corrosion. Polymer optical fibers are also much more elastic and can be stretched by about 20% without losing their elasticity. Problems can occur with polymer fibers when chemical interactions arise with the medium to be measured or in oil industrial applications with glass fibers because standard glass fibers are sensitive to hydrogen and tend to increase its attenuation. Special protective coatings are necessary.

If extended structure components have to be evaluated, another great advantage is their ability to perform long-range measurements over distances up to several tens of kilometers without the need for any electrically active component, whereas any measuring effect can be identified everywhere along the optical fiber (distributed sensing).

This short overview is not intended to give a comprehensive insight and detailed description of this sensor technology. It should rather give an idea of unique possibilities to investigate complex materials and to explore the behavior of materials by embedded tiny and highly sensitive sensors.

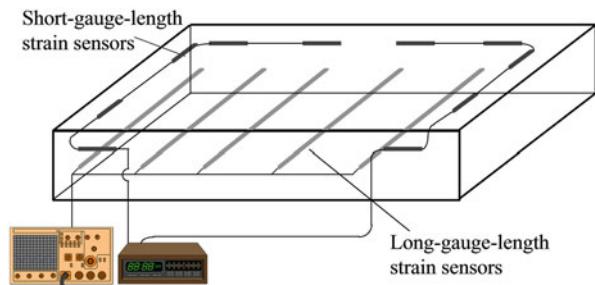
---

## Optical Fiber Sensor Classes and Sensing Technologies

There are many different sensor types and physical principles used for monitoring purposes (see ► Chap. 44, “Structural Health Monitoring”) and for investigation of the material’s behavior. We can basically distinguish between two categories: long-gage-length sensors with gage lengths above the meter range and short-gage-length sensors with gage lengths in the millimeter and centimeter range (see Fig. 1). Optical fiber sensors (OFS) are widely established to monitor the behavior of large structure components and extended structures and buildings because of the specific feature that long fibers can bridge measuring zones in the meter range (e.g., integral deformation measurement) and/or detect material changes along the whole fiber length with the capability to identify the position where the measurand is acting.

The complement to long sensor fibers is short-gage-length sensors. Such local optical fiber sensors (sometimes referred to as point sensors) have very small gage lengths in the millimeter range (5–20 mm). In the case of strain sensors, they are comparable with resistance strain gages. Such local optical fiber

**Fig. 1** Optical fiber sensor categories (embedded or attached to the surface) as an integrated part of a smart structure

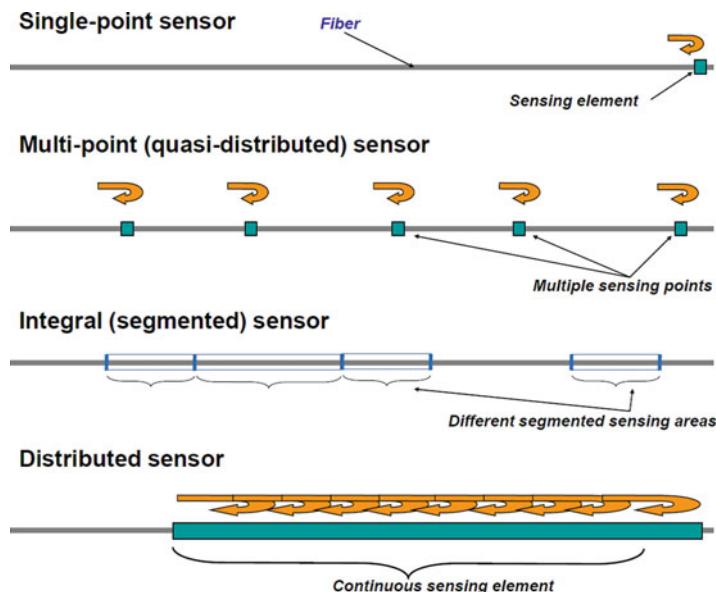


sensors can also be used as pH sensors or humidity sensors (see below). For NDE of materials, local OFS are embedded into materials and preferably used to detect the long-term and/or load-depending behavior in highly stressed zones of composite or laminated materials. In the same way, extremely tiny optical fiber sensors can be embedded into curing plastic or mineral materials to detect the deformation behavior during the curing process. Specially designed fiber sensors allow measurement of deformation already in a very early curing state, for example, in the case of mortar, concrete, or plaster in the interphase between rheology and solid state because such a sensor does not constrain the surrounding non-stiff matrix from deformation.

In spite of these unique features, short-gage-length fiber sensors are confined to rather low-volume niche applications where they remain the only possible choice. Section “[Application Examples](#)” will show typical application examples where other sensor technologies cannot be used or cause some difficulties in interpretation of the measurement results.

Summarizing the considerations above, optical fiber sensors can be classified into four groups with regard to their gage lengths (see Fig. 2):

- Single-point sensors:  
The sensitive area along the optical fiber is localized to a very small area; a discrete sensor measures at a predetermined point (point sensor); the point sensor has a very small gage length (similar to resistance strain gages) and can be positioned at the end of the fiber or along the fiber at any place.
- Multi-point sensors:  
Single sensitive areas are available along the optical fiber; several single-point sensors can be read out successively or simultaneously. This multi-point sensor arrangement, for example, a fiber Bragg grating chain of several single sensors, is also called “quasi-distributed” sensor.
- Integral and/or segmented sensors:  
The sensitive area has a defined gage length, i.e., the sensor has a limited spatial distribution of a few centimeters up to a few meters, and provides integral information over a defined gage length between two defined points along the optical fiber.
- Distributed sensors (also called fully distributed sensors):  
The magnitude of a physical (or chemical) parameter or a measurand and its variation can be measured along the entire length of a continuous uninterrupted



**Fig. 2** Optical fiber sensor types with regard to the gage length

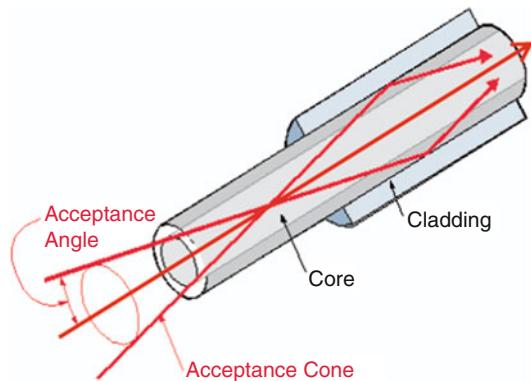
optical fiber. Distributed sensors are based on backscattering effects, such as linear (Rayleigh) backscattering and nonlinear (Raman and Brillouin) backscattering (see also ► Chaps. 11, “Surface Brillouin Scattering” and ► 17, “Raman Scattering”). Because the scattered signal is recorded against the time, the location of the measuring event (measurand’s change) can be determined within a specified spatial resolution, which means the minimum distance over which the system is able to indicate the value of the measurand.

All different sensor types are explained and selected examples are described in the following sections.

Concerning optical fiber sensing technologies in use, a few words about the physical principle are in order. When light is transmitted into one end of the fiber surrounded by an optical cladding with a lower refractive index than the core, it propagates through the fiber to the other end due to the physical effect of total internal reflection. Depending on the accepted angle of incidence at the core-cladding interface, and depending on the diameter of the optical fiber core, the propagating wave packets are more or less reflected at the core-cladding interface. (If the angle of incidence is higher than accepted for total reflection conditions, light leaves the fiber core and is usually not useful.) Figure 3 shows this effect.

If the transmission of the light wave is affected by external influences (measurand to be detected), one or more physical parameters of the light wave in the optical fiber are varied (sensor). This variation is detected, recorded, processed, and must be re-transformed into a scaling unit of the measured quantity. The great challenge for

**Fig. 3** Light beam propagation in an optical fiber  
(Website “Fiber”)



the engineer is to separate the variations induced by the measuring object from any variations induced by some other internal or external effects. Often faulty measurements are produced by an inappropriate application of the sensing element.

In order to make an optical fiber into a sensor, one or more physical parameters of the propagating wave have to be varied by the measurand. All known parameters like intensity, wavelength or phase of the light wave, and its polarization state or a wavelength shift (color change) can be modified by a measurand. To determine a position in a fiber, the flight time of a light pulse transmitted into the fiber end and back-reflected at the position of impact is measured. If the flight time to defined markers in the fiber is observed, for example, in the case of segmented fiber sensors, the shortening or extension of the optical path length (contraction or extension of the sensor fiber) can be assessed from measurement of the time of transit. It can be stated that a considerable number of optical fiber sensor types have been created in the past decade for measurement of almost all physical and a lot of chemical quantities. This advancing development of optical fiber sensors is not insignificantly affected by the development of components and tools for optical communication and data links.

On the one hand, it is important to discriminate between transmission effects where the light is guided in the fiber and its physical parameters like intensity, phase, or velocity are affected by the measurand and, on the other hand, backscattering effects, where the launched light (laser) beam interacts with the constituent atoms and molecules of the fiber material. These backscattering effects allow unique distributed monitoring with applied optical fibers to evaluate the integrity or detect damaged zones of extended structures and/or to measure strain, temperature, vibrations, and also moisture over extended areas. Details are described later in this chapter.

A large number of optical fiber sensors in different designs using all possible optical effects allow their application for the measurement of strain, displacement, pressure, vibration, acoustic emissions, temperature, determination of the position of damage along an installed optical fiber, and even the detection of partial discharges inside high-voltage isolators. If optical fibers are coated with chemically sensitive layers, detection of chemical and biological species, early

detection of beginning steel corrosion in reinforced concrete, as well as the detection of harmful chloride concentrations in prestressed concrete components are possible.

One of the most recognized advantages of optical fiber sensors is that all parts of the sensing components are immune to high voltage (HV) and electromagnetic interference since their components are made from exceptional dielectric materials. They also do not need any electrical power supply in the sensing area and provide complete galvanic separation between high-voltage potential and any devices operating on earth potential. Their tiny size and compactness enable more or less intimate integration into facilities of high-voltage substations, power transformers, generators, and also into the components of HV cable lines. These advantages push numerous activities to monitor and evaluate the operation as well as the condition state and/or integrity of the HV equipment. Optical fiber sensors are therefore directly embedded in the insulating material on high electric potential systems to measure and evaluate, among other quantities, temperatures, vibrations, partial discharge activities, water penetration, and chemical analysis of gas-insulated switchgears outside of their regular behavior.

## Sensors for Local Measurements

### Fiber Bragg Grating Sensors

A number of NDE tasks focus on observation of the long-term behavior of newly developed or quite complex materials such as high-performance concrete, modified mortars for historic structures, composite materials for aeronautics or rotor blades of wind turbines, and combined materials such as plastic-covered concrete structures. In order to perform continuous observation or to detect dramatic changes of the structure component's integrity, embedding of tiny sensors including leading cables is recommended. If the area to be observed is limited, local sensors – so-called single-point sensors – are preferred.

The best known short-gage-length optical fiber sensor is the fiber Bragg grating (FBG) sensor. FBG sensors are inscribed into the core of an optical fiber and can therefore be bought as a fiber length with one or more fiber Bragg gratings inscribed into the fiber core. FBG sensors allow measuring strain and temperature. It should be noted that both measurands are simultaneously recorded and have to be separated if only strain or temperature is being measured.

A fiber Bragg grating strain sensor is very similar to the widely known metallic wire-type (resistance) strain gage. In the same way as the resistance strain gage consists of a sensitive element (metallic zigzag foil pattern) and a supporting material onto which the conductive strip is fixed, FBGs can be attached to or integrated into supporting materials. Figure 4 shows examples of FBGs. The design of such prefabricated FBG sensor patches depends on the measurement task and the material of the measuring object. They are available either as a single sensor or as a two-dimensional strain gage rosette. Such prefabricated FBG strain sensors can be glued or spot-welded to structure components and materials to measure strain in one,



**Fig. 4** FBG strain gages (patches), left: for one-dimensional and two-dimensional measurements ([www.hbm.com](http://www.hbm.com)), middle: with ready-to-connect cables ([www.micronoptics.com](http://www.micronoptics.com)), right: specially coated sensor fiber OptiMet (HBM GmbH, Germany) for use under harsh environmental conditions

two, or three directions. If only strain in one direction is to be measured, the optical fiber can be glued directly to the measuring object (see Fig. 17).

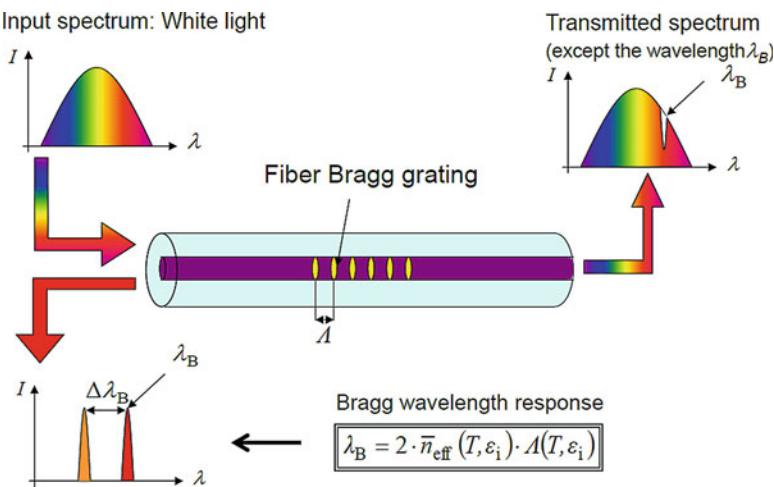
Fiber Bragg gratings are phase diffraction gratings inscribed into optical waveguides by using UV light (e.g., by an excimer laser at 248 nm). The fiber is exposed to an interference pattern of this UV radiation. UV photosensitive processes produce changes in the refractive index of the fiber core which is susceptible to these. The interference pattern is an image in the fiber core of a periodically changing refractive index. This periodicity in refractive index defines the FBG period  $\Lambda$ . Incident and transported light along the fiber is additively superposed for a certain wavelength in this interference pattern area (constructive interference). Figure 5 illustrates the operating principle. When broadband light (input spectrum) travels through the grating with periodic alterations in the fiber core's index of refraction, only one sharp wavelength (so-called Bragg wavelength  $\lambda_B$ , red arrow, left, in Fig. 5) of the white light spectrum is reflected. The amount of  $\lambda_B$  depends on two factors: the index of refraction  $n$  and the distance of the grating planes  $\Lambda$ . If the grating structure is longitudinally deformed either by mechanical or by temperature influence (or by both), the distance  $\Lambda$  or the index of refraction  $n$  is changed, and, as a consequence, the Bragg wavelength  $\lambda_B$  is shifted.

The physical dependence on  $\Lambda$  and  $n$  can be described by the general equation:

$$\begin{aligned}\Delta\lambda_B = & 2 \cdot \left( \Lambda \frac{\partial n_{\text{eff}}}{\partial L} + n_{\text{eff}} \frac{\partial \Lambda}{\partial L} \right) \cdot \Delta L \\ & + 2 \cdot \left( \Lambda \frac{\partial n_{\text{eff}}}{\partial T} + n_{\text{eff}} \frac{\partial \Lambda}{\partial T} \right) \cdot \Delta T\end{aligned}\quad (1)$$

The first term in Eq. (1) describes the effect resulting from the mechanical deformation ( $\partial\Lambda/\partial L$ ) and the elasto-optical reaction ( $\partial n_{\text{eff}}/\partial L$ ) of the optical waveguide; the mechanical deformation could be expressed alternatively as  $\varepsilon_z$  instead of  $L$ . The second term in Eq. (1) describes the temperature effect on the quantities  $n_{\text{eff}}$  and  $\Lambda$ . The term ( $\partial\Lambda/\partial T$ ) describes the effect of the thermal expansion of the Bragg grating on the grating period  $\Lambda$ . The thermal effect on the refractive index of the optical fiber, on the other hand, is expressed by the term ( $\partial n_{\text{eff}}/\partial T$ ).

Assuming uniform axial strain changes in the grating area and the absence of lateral deformation of the grating, the strain seen by a grating can be computed by a simple linear equation:



**Fig. 5** Schematic description of the Bragg grating sensor function. One narrow pulse of the broadband light traveling through the grating, the so-called Bragg grating wavelength  $\lambda_B$ , is reflected back to the receiver. If the grating period  $\Lambda$  changes due to deformation of the grating or if there is a variation of  $n$  due to temperature variations, the Bragg wavelength moves in the spectrum. This wavelength change  $\Delta\lambda_B$  represents the measurand and is read out

$$\varepsilon = K \cdot \frac{\Delta\lambda_B(\varepsilon_z)}{\lambda_B} + \xi \cdot \Delta T \quad (2)$$

$K$  and  $\xi$  have to be estimated by a calibration procedure; recommendations on how to perform calibration are given in IEC standards 61757-1-1 and 61757-2-1.

In order to measure the shift in the Bragg wavelength  $\lambda_B$  caused by strain and/or temperature changes, two other system components are required: an interrogator device and a processing unit, which takes recorded data from the interrogator and performs further data processing, transmission, and storage. The interrogator illuminates one sensor or a chain of sensors and records the reflected FBG signal with the wavelength  $\lambda_B$ .

If there is a need to resolve strain changes of  $1 \mu\text{m}/\text{m}$  or temperature changes of  $0.1 K$ , the measurement of the Bragg wavelength  $\lambda_B$  must be better than 1 part in a million. For laboratory investigations, commercially available optical spectrum analyzers (OSA) are used. They are expensive, bulky, and usually fragile, but they enable detailed evaluation of the reflected Bragg wavelength and its change including the spectral width and other characteristic features of the Bragg pulse. Some suppliers provide OSA for measurement in the field with lower resolution. The scanning principles are based on curve fitting or on the use of a centroid algorithm. The disadvantage of such devices is the expense of the CCD array operating at wavelengths above  $1 \mu\text{m}$  and the moderate data rate achievable. Some of the disadvantages are being overcome by sophisticated solutions. Currently, compact models are offered, which operate up to 35 kHz (**I-MON Interrogation Monitors**).

For on-site measurements, different signal processing schemes have been developed. Depending on accuracy and reproducibility requirements, a suitable interrogation system should be chosen. In the following, the main types of interrogation techniques are mentioned.

One can be distinguished between highly sophisticated and less sophisticated interrogation systems through the use of a broadband light source or the use of a laser source. Devices using a broadband light source are coupled with different kinds of spectral filters to measure the reflected Bragg wavelength. One group of the filter-based devices uses Fabry-Perot filters and edge filters. Fabry-Perot interferometer-based interrogators result in the highest optical resolution of all the interferometers.

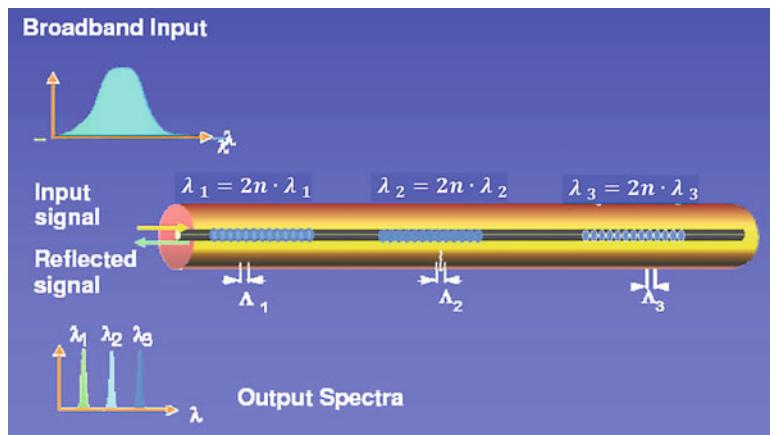
A scanning Fabry-Perot filter (SFFP) consists of two closely spaced highly reflective flat mirrors, one of which can be moved along the optical axis using a piezoelectric actuator. The wavelength shifts are converted into phase shifts, which are detected by variations in the light intensity of the (nonlinear) interference signal due to the varying path differences (moving mirror). The advantage is that resolutions of about 0.01 nm in wavelength changes are possible; however, such devices are rather expensive and prone to environmental interference.

Another interrogator type uses a sloped optical filter, in which a shift in the grating wavelength is converted directly into an intensity change. The light reflected back by the grating passes through a filter with a known pass/reject ratio, i.e., the transmission is a known linear function of wavelength. The narrowband reflection from the (measuring) grating can then be determined simply by measuring and comparing the passed and rejected intensities. This simple way of reading the FBG wavelength change results in inexpensive devices, but the disadvantage is that it can address only one grating at a time.

Alternatively, tunable or swept-wavelength light sources can be used together with simple photodiode detectors. A scan generator tunes the light source, sweeping it through the wavelength space of the sensor system such that at any given instant the wavelength of light being transmitted down the fibers is known. When this wavelength coincides with the Bragg wavelength of an FBG, light is reflected back to a photodetector. The Bragg wavelength to be measured can be easily identified.

FBG sensors offer multiplexing capability due to the wavelength-encoded nature of gratings. Each grating in a fiber can be designed to have its own wavelength in such a way that the distance  $\Lambda$  of the grating planes is different. Different grating designs lead to different Bragg wavelengths within the available wavelength spectrum of the light source. Figure 6 shows an optical fiber with three FBG sensors (FBG array), which produce separate wavelength responses. If the measurand only acts on one of the grating, the corresponding Bragg wavelength change can be recorded separately. This enables multi-point sensing and thus recording of strain or temperature profiles in a possibly extended structure with only one sensor fiber containing a number of FBG sensors. If wavelength multiplexing is applied, up to about 20 FBG sensors can be read out by one device, depending on the load-induced shift of the wavelengths.

In order to read out several gratings in one fiber, interrogators with demultiplexing capability are needed. FBG sensors described in the previous section



**Fig. 6** FBG array with three separately measuring FBG sensors. Each grating has its own grating period  $\Lambda_1$ ,  $\Lambda_2$ , and  $\Lambda_3$ , which generates the Bragg wavelengths  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$ . These wavelengths and their changes can be detected and measured separately and allow measuring, for example, strain profiles along a measuring object

were characterized by different distances  $\Lambda$  of the grating planes, i.e., of different Bragg wavelengths. Interrogation techniques that use a broadband source can interrogate multiple sensors of different wavelengths on a single fiber. To read out a moderate number of FBG sensors in one fiber, Fabry-Perot filter or spectrometer-based systems can be used. The number of possible FBGs is limited by the bandwidth of the broadband source and by the expected measuring range, which defines the maximum wavelength changes in positive and negative directions. For example, if the bandwidth of the broadband source is 40 nm (typical value in the range of 1,550 nm) and the needed wavelength range is 4 nm to cover the wavelength changes during operation of the sensor, ten sensors can be read out without the danger of overlapping the pulses moving in the optical spectrum. This wavelength-division multiplexing (WDM) method is widely used.

If significantly more FBGs in one fiber have to be read out, time-division multiplexing (TDM) approaches are used. This requires different gratings of the same or similar wavelengths. In this case, the time for their return signals to reach a detector (the return time from closer gratings is shorter than those from more distant ones) is measured by a pulsed broadband light source. Alternatively a high-speed spectrometer can be used. There are some disadvantages of TDM systems. To read out many FBGs, they must have low reflectivities to ensure sufficient pulse power available from each grating. The spacing between the gratings must be large enough to allow the interrogator enough time to make a wavelength measurement and to distinguish between the return times from the gratings; when reading out tens of gratings, the grating separation should be 1 m or more to avoid cross talk. And finally, an important disadvantage is that in case of a fiber break close to the interrogator, the whole sensor chain is lost. These disadvantages limit the performance and practicality of TDM measurement systems ([Smart Fibres](#)).

Depending on the interrogator type and the number of simultaneously recording sensor fiber channels, the upper frequency limit of commercial devices is in the range of 5 kHz for several channels and up to 200 kHz for a one-channel device. Special methods are being developed to increase the scan frequency to enable the use of FBG sensors for detection of dynamic such as acoustic emission or impact events.

Another important aspect especially when long-term or very accurate measurements have to be carried out is the reference to a stable value, which is not influenced by aging effects or by environmental influences. High-precision interrogators use an internal NIST-traceable gas cell to ensure accurate measurements on every sweep of the powerful laser ([Micron Optics sm125 Interrogator](#)).

It must not be forgotten that, if both strain and temperature influences exist, a reference grating which only senses the temperature has to be used to separate the strain value from the temperature value. Vice versa, the deformation influence must be eliminated to measure temperatures with a FBG temperature sensor ([VDI/VDE 2660](#), part 2). If reliable data recording is needed, strain and temperature resolutions of about 1  $\mu\text{m}/\text{m}$  or better or 0.1 K or better, respectively, are possible.

FBG sensor is also delivered as an optical fiber mostly indistinguishable from common communication fibers. The optical fiber with a diameter of about 200  $\mu\text{m}$  contains one or more FBG sensors; its length is of the order of 10 mm. Such tiny fibers with inscribed strain sensors are often embedded into structure materials or applied to surfaces. There are many constructive versions how FBG sensors are designed and applied. The FBG sensor can be embedded in a protective material which constitutes a transition zone between the sensor element and the object of measurement. The transition zone is usually flat or planar. A FBG sensor can also be used similar to an extensometer where the connecting fibers of the FBG sensor are fixed to the object of measurement at anchoring surfaces/points of defined distance. An alternative (extensometer) version is its installation inside a tube; the fiber is then strained (strain sensor) or unstrained, i.e., loose (temperature sensor). Selected examples for FBG sensors are shown in section “[Sensors for Local and Quasi-distributed Measurements](#).”

There are many different strain sensor patches adapted to special application and environmental conditions. Figure 7 gives an impression on the availability.



**Fig. 7** Commercial FBG strain sensors. Left: HBM GmbH, Germany (<https://www.hbm.com/en/>), middle: Micron Optics, USA (<http://www.micronoptics.com/>), right: Stran Technologies, USA (<http://www.strantech.com/>)

If FBGs are installed free of external forces which cause mechanical deformation, only the thermal influence takes effect and temperatures can be measured. FBG can be installed in small tubes, as surface-applied or embedded temperature sensors. They can also be positioned in liquids or gases and can work up to 700 °C. Figure 8 shows examples of commercially available FBG temperature sensors.

Coatings of FBG sensors can be designed to be sensitive to other physical quantities or chemical species. If the coating consists of a moisture-sensitive polymer film, e.g., polyimide coating, the Bragg wavelength of the polyimide-coated FBG changes while it is exposed to different humidity conditions (see Fig. 9, left). The wavelength shift is caused by the volume expansion of the polyimide coating. Such sensors can work in the relative humidity (RH) range from about 11% RH to about 97% RH. However, it is easier to measure gas humidity than material humidity with regard to the calibration (Venugopalan et al. 2009).

Another important physical quantity is pressure which often has to be measured under strong electromagnetic influences or even in explosive environment. FBGs with its electromagnetic immunity are either integrated into materials to detect pressure, or special pressure probes are designed, in which FBGs are deformed. Because the intrinsic (lateral) pressure sensitivity of FBG is not high, those sensors are mainly designed to detect pressure changes by indirect measurement of axial strain in the FBG. A very common approach is attaching the FBG to a flexible diaphragm in the diaphragm plane where the strain is expected to be maximal. Such



**Fig. 8** Commercial FBG temperature sensors. Left: AOS GmbH, Germany (<http://www.aos-fiber.com/>), middle: FBG KOREA, Inc., South Korea (<http://fbg.co.kr/eng/>), right: O/E Land Inc., Canada (<http://o-eland.com/>)



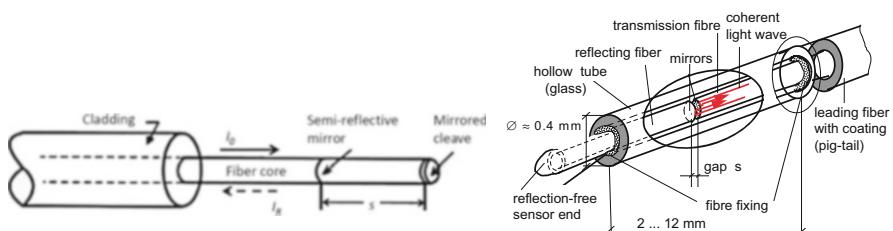
**Fig. 9** Commercial FBG sensors for physical measurands. Left: Humidity sensors (O/E Land Inc. Canada), right: FBG pressure sensor (Safibra, s.r.o., Czech Republic) for pressure up to 150 bar

a design leads to a bulky sensor and is rather used in the high pressure range. Figure 9 (left) shows a FBG pressure sensor for applications in civil engineering as well as in the oil and gas industry where a bigger sensor size is acceptable. If the sensor size has to be minimized, FBG sensors can be mounted in small cylindrical tubes with an outer diameter of 1 mm or less. The grating is deformed due to mechanical pressure; however, in case of hydrostatic pressure, the sensitivity is low. Due to the known sensitivity of FBG to temperature, there are some practical problems, if the temperature is not constant. Temperature compensation is necessary, usually done by two sensing FBGs. One FBG senses pressure, and the second FBG is used for parasitic temperature calibration. Commercial FBG pressure sensors are used to observe process parameters and to evaluate the integrity of pipelines, valves, and vessels. They can also be applied to porous stones so that groundwater or seawater pressure can be measured.

### Fiber Fabry-Perot Sensors

If there are higher-frequency and measurement value resolution requirements, another short-gage-length optical fiber sensor is used: the Fabry-Perot interferometer (FPI) sensor. Interferometric sensor principles generally allow reaching very high measurement resolutions (see ► Chap. 12, “Interferometric Methods in NDE”), in the case of metric quantities down to the range of nanometers and/or strain resolution of  $<0.1 \mu\text{m/m}$ . Although the evaluation of the periodic signal response is more challenging and there is no capability of multiplexing, there are a few advantages in the field of NDE of the materials’ microstructure. For this reason, the basics and the usefulness of FPI sensors are considered here.

To produce an interference signal, a monochromatic beam (e.g., a laser beam) must be reflected twice on its way, and the reflected phase-shifted parts of the initial beam can interfere. This can be done by creating reflectors at locations in the length of a simple optical fiber (see Fig. 10, left). These two mirrors form a cavity of, for example, a few millimeters, where the sensor signal is produced. An externally acting measurand, which produces force-induced or temperature-induced deformation of the fiber, changes the length of the cavity and induces phase changes in the optical signal. These phase changes are detected as an intensity change in the output interference signal recorded (compare to Eq. (3)). In this sensor type, the cavity length defines the gage length of the sensor, which must be known to calculate stress values from the measured deformation values.

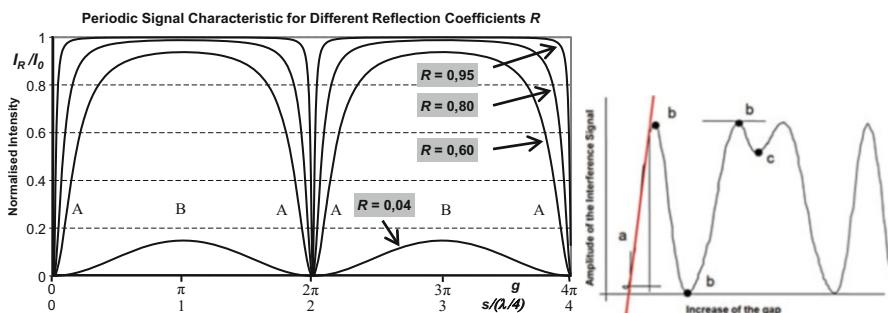


**Fig. 10** Types of fiber Fabry-Perot sensors. Left: intrinsic type, right: extrinsic type

A more common design is the extrinsic fiber FPI; it is easier to manufacture and enables several other advantages. In this case, the two mirrors are end faces of two optical fibers, where one fiber end is positioned opposite to the other fiber end face forming a small gap of usually some  $10\text{ }\mu\text{m}$  up to a few millimeters. Such an extrinsic FPI sensor (also called EFPI) is shown in Fig. 10, right. To hold the two fibers in position, a hollow glass or steel tube with an inner diameter of a little more than the fiber diameter (about  $130\text{ }\mu\text{m}$ ) can be used. This design forms the cavity, usually in air, where interference is initiated. If the cavity must be sealed from external influences, both fiber pieces (input/output fiber and the dead fiber end) are fixed at the ends of the tube. The measuring effect is induced when the hollow tube is axially deformed. The two fiber end faces move inside the tube in its axial direction with the result of a change of the cavity length (gap width  $s$ ). This produces a phase difference between the two beams reflected at the fiber end faces (see Fig. 10) and changes periodically the backscattered recorded intensity  $I_0$  of the interference signal (interference fringe function):

$$\frac{I_R}{I_0} = 2R \left[ 1 + \cos \left( \frac{4\pi s}{\lambda_0} \right) \right] \quad (3)$$

where the reflectivities  $R_1$  and  $R_2$  of both fiber end faces are assumed equal ( $R_1 = R_2 = R$ ). Equation (3) describes the typical output signal  $I_R$  depending on the changing length of the cavity  $s$ .  $I_0$  is the intensity of the launched beam. The periodic function can be exploited in two ways. If the requirement on resolution is moderate, the interference fringes, i.e., the cycles between the crest values (light-dark transitions of the interference fringes), are counted. The resolution is then related to a quarter of the wavelength  $\lambda$  (compare Fig. 11). If the resolution requirement is very high, or if very small deformation changes must be measured, the intensity variation along the rising or falling edges of the cosine function (range A in Fig. 11, left) is evaluated. The higher the reflectivities at the end faces of the fibers are, the steeper is the interference function, also called finesse.



**Fig. 11** Use of the interference function of a fiber Fabry-Perot sensor for measurements with very high resolution. Left: Slope of the periodic function depending on the reflectivity  $R$  of the fiber end faces ( $s$  – cavity length, i.e., deformation of the sensor;  $g$  – phase increment). Right: Exploitation of the linearized part of the interference function to reach maximum sensitivity

Considering that the function is linear between the crest values  $b$  of the cosine function (see Fig. 11, right), intensity changes can be measured very precisely. The better the discretization of the A/D converter is, the higher is the resolution of deformation changes. Higher demand on data processing is also necessary, if end face displacements in the millimeter range are to be measured. Depending on resolution and measurement range (gage length) requirements, a combination of fringe counting and resolution increase of the intensity changes is necessary. Another problem is the decreasing sensitivity near the crest values of the periodic function; it becomes zero at the maxima and minima values  $b$ ; point  $c$  shows a direction change in deformation. More details can be found in Habel (2004).

One advantage of EFPI sensors is again their diameter, which is less than 1 mm. The design, shown in Fig. 10, right, where the reflecting fiber pieces are fixed at both ends of the capillary, is almost insensitive to temperature changes because any thermally induced deformation of the components (glass capillary and fibers) is compensated. Another unbeatable advantage of EFPI sensors can be achieved by a design modification. A modified EFPI sensor described later (see Fig. 23) enables one to measure deformation without producing significant shear stress at the interface between the sensor and the surrounding measurement zone. Such a very tiny sensor design enables displacement measurements inside, for example, mineral or epoxy materials during curing or in soft materials. As a compliant fiber Fabry-Perot deformation sensor, it can react down to an applied axial force of about 150  $\mu\text{N}$ .

Very tiny EFPI sensors with glass or metal capillary are specially manufactured for laboratory investigations. Specific skills are necessary to use such highly sensitive sensors on construction sites. If robust interferometer-based extensometers or pressure sensors or accelerometers are required, commercial Fabry-Perot interferometer sensors with high resolution of the measured values are available in different designs adapted to several measurement tasks, physical quantities, and environmental conditions. Figure 12 shows examples of deformation and pressure sensors. They are highly statically as well as dynamically sensitive; however, they do not allow



**Fig. 12** Examples of commercially available fiber Fabry-Perot interferometer-based extensometer (left), strain gage (middle), and pore-water/fluid pressure sensor (right) ([www.roctest.com](http://www.roctest.com)). FPI sensors enable measurements with high accuracy. Such sensors are not affected by electromagnetic waves or lightning strikes and can be installed in areas or plants with potentially explosive atmospheres

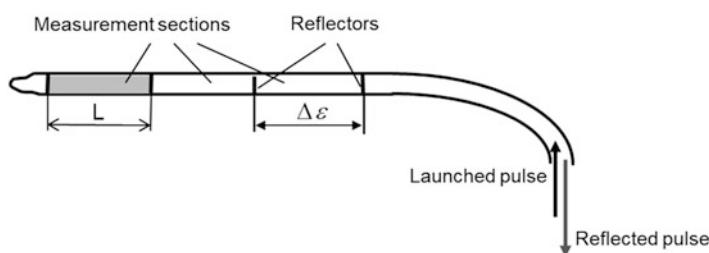
measuring deformation of soft or curing materials, for example, in the interphase between rheology and solid state.

If several interferometer sensors are needed to measure at various places simultaneously, they can be scanned by switching from one sensor to the next. If optical fiber sensors with minimum size and extremely low stiffness are needed, specific designs of movable (non-stiff) extrinsic Fabry-Perot interferometers were developed. They enable investigating and evaluating the microstructure of mineral materials, composite elements, layers, etc. without disturbing the zone to be investigated. Examples are described in section “[Sensors for Local and Quasi-distributed Measurements](#).”

### Multiplexed Single-Point and Segmented Sensors

One of the greatest advantages of optical sensors based on waveguides is that several measurement sections can be generated along one single optical fiber. There are basically two options: first, positioning several sensors with multiplexing capability along the fiber (Fig. 6) and, second, creating measurement sections with defined gage length(s) along the fiber (Fig. 13) where deformation and/or strain profiles can be detected. The position of the reflectors is determined by launching a short light pulse into the fiber and measuring the time of flight of the back-reflected signal. This historically well-known methodology comes from telecommunication applications, where the position of damage along data communication lines can be found using this methodology – the optical time-domain reflectometry (OTDR). In sensing cables, it is used to indicate the position of specifically positioned reflectors and its possible change in the fiber to evaluate an acting mechanical measurand.

In such segmented sensing fibers, the distance of a reflector in the fiber is proportional to the flight time of a laser pulse there and back to the photodetector. Comparing the flight time of two neighboring reflectors, their distance can be defined. A change in the pulse propagation time  $\Delta t$  can be interpreted as a shift of the reflectors  $\Delta l$ , i.e., an increase in the time difference of two neighboring reflectors



**Fig. 13** Quasi-distributed optical fiber sensor with different measurement sections. The optical fiber contains reflectors which reflect a part of the launched laser pulse back to the recording device. Measurement of the flight time of the laser pulse allows calculation of the position of the reflectors and, correspondingly, distance changes between several reflectors if they move. Movement of reflectors represents deformation of the material to be observed

can be interpreted as an increase of the measuring segment due to mechanical loads as well as, vice versa, in the case of shortening the flight time as a decrease of the observed extension of the measuring object. The proportionality coefficient  $K$  between the change in length and time depends on the phase velocity of light in the optical fiber  $c/n$  and on the photoelastic coefficient  $a$ :

$$K = \frac{\Delta l}{\Delta t} = \frac{c}{2n} \cdot \left( \frac{1}{1+a} \right) \quad (4)$$

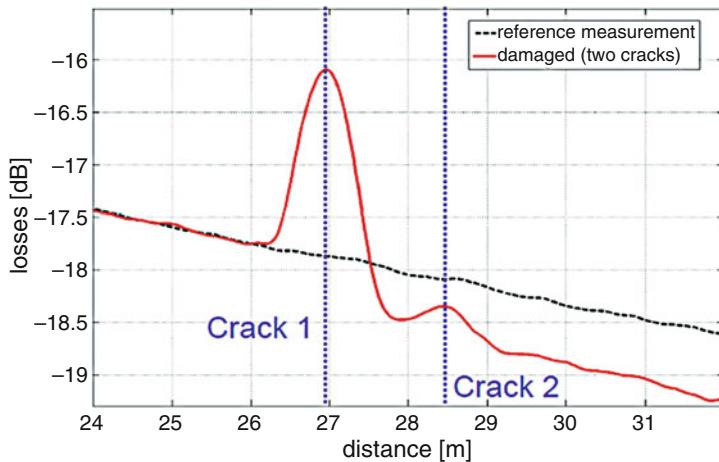
where  $n$  is the group refractive index of the fiber,  $c$  is the speed of light in vacuum, and  $a$  is the photoelastic coefficient. It considers changes in the relevant optical glass parameters, when the fiber is strained. In order to evaluate the influence of photoelasticity, the fiber sensor must be calibrated in the laboratory before applying it on-site. The refractive index of the optical fiber is usually known.

The areas to be evaluated in both so-called segmented sensor arrangements can be read out separately so that changes in the measurement quantity can be localized along the whole fiber length. Examples of FBG-based strain and/or temperature profile sensors as well as segmented deformation sensors are described in section “[Sensors for Local and Quasi-distributed Measurements](#).”

## Sensors for Distributed Measurements

One unique feature of optical fiber sensors is their capability of sensing over the entire fiber length (distributed sensing). An optical fiber can be used to measure physical or mechanical property changes at any position along the fiber with over hundreds of meters or even several decades of kilometers. The minimum distance of distributed sensors over which the sensor system is able to indicate the value of the measurand within the specified uncertainty (spatial resolution) depends on the quality of the measurement instrument and is of the order of a few centimeters up to several meters. Application of such methodologies for NDE purposes is recommended for observation of extended structure components. Basic terms of optical fiber sensors can be found in guidelines, for example, Guideline for Use of Fiber Optic Sensors (2009), IEC 61757:2018, and ASTM F3092-14 (2014).

If an optical pulse wave launched into the fiber only interacts with local macroscopic or microscopic density fluctuations in the fiber, the scattered light is sent back with the wavelength of the incident light. This elastic scattering mechanism is called **Rayleigh scattering**. An unaffected optical fiber shows a continuously decreasing intensity according to the transmission loss of the fiber material, represented by the dotted line in graph (Fig. 14). If the fiber is locally strained, mechanically deformed, or physically influenced, a change in the attenuation profile shows this perturbation. Depending on the intensity of the defect (crack, strain, and fiber deformation), the intensity of the peak varies. To determine the position of such a perturbation, the time of flight of the backscattered signal is measured (OTDR methodology). This feature of optical fibers can be used for nondestructive evaluation of the behavior of



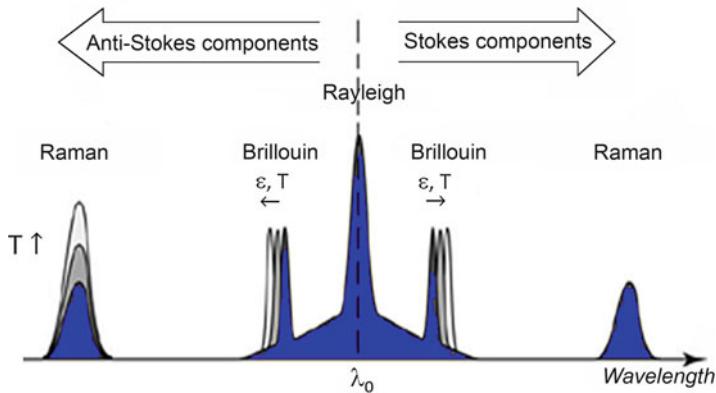
**Fig. 14** Typical detection of an attenuation change in a strained polymer optical fiber caused by cracks in the material to be evaluated; the position of cracks is recorded by OTDR (Liehr et al. 2008)

extended structures as well as for early damage detection. Depending on the expected deformation, optical polymer or glass fibers can be used. In case of distributed polymer fiber sensors, the length of detection is limited to about 500 m.

Alternatively, distributed optical fiber sensing techniques are also based on non-linear optical effects, where the light wave interacts with the matter. Interaction leads to an energy exchange and is associated with frequency shift of the back-reflected light. There are two such inelastic backscattering effects exploited for distributed sensing: Raman backscattering mechanism exploited for distributed temperature sensing and Brillouin backscattering mechanism exploited for distributed strain and also temperature sensing. Figure 15 summarizes schematically the three backscattering mechanisms. The center of the graph shows the Rayleigh backscattering signal back-reflected with the same wavelength (elastic scattering effect). Back-reflected waves with different wavelengths show the Raman signal and the Brillouin signal. The downshifted so-called anti-stokes wavelength line is clearly temperature dependent and used in Raman-based distributed fiber-optic temperature sensor systems. In Brillouin-based distributed sensors, the value of the strain or the temperature is estimated using the shift of the peak wavelength/frequency, so-called Brillouin frequency.

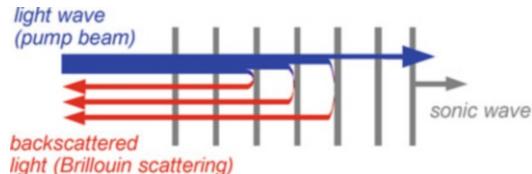
**Raman-based distributed sensor systems** are used to control varying temperature profiles in extended and complex structure materials such as huge concrete dams and new composite materials under climatic influences to evaluate damaging mechanical stress effects. The Raman-scattered light is caused by thermally excited molecular vibrations. Another possibility is to observe temperature distribution in the oil and gas industry, for example, leakage detection in oil wells, along pipelines, or in underground gas storage facilities using the Joule-Thomson effect (Grosswig et al. 2001).

**Brillouin-based distributed sensor systems** are used to evaluate the mechanical integrity of extended structures like long bridge girders, blades of wind turbines, pipelines for oil and gas, dams and river dykes, rock areas, and slopes. This inelastic



**Fig. 15** Backscattering effects used for distributed optical fiber sensing.  $\lambda_0$  is the wavelength of the incident light

**Fig. 16** Principle of stimulated Brillouin scattering. (Courtesy: Dr. Katerina Krebber, BAM)



scattering effect can be stimulated by mechanical excitation of the fiber glass, e.g., by a strong laser pulse. A strong laser pulse (also called pump pulse) launched into the end of one fiber – called pump beam – induces acoustic vibrations (so-called phonons) in the material. Such sound waves can be considered as periodic, mechanical density fluctuations in the fiber. These acoustic waves meet the scattered low-intensity optical wave (photons), and the resulting backscattered signal is recorded. Figure 16 shows this mechanism schematically. Because the phonons are provided by excitation, the physical effect is called stimulated Brillouin scattering (SBS). A stimulation of the Brillouin scattering process occurs when the frequency difference or wavelength separation of the pump pulse and the probe signal corresponds to the Brillouin shift (resonance condition); both optical signals are counterpropagating in the fiber. Due to superposition of the electric fields of the high-energy pump and the scattered optical wave, a beat with the frequency  $\Delta\nu/2$  propagates with the velocity  $v_a = c/n \odot \Delta\nu/\nu_0$ , where  $\nu_0$  is the mean frequency of light waves and  $c/n$  the speed of light in the waveguide.

Exploiting this inelastic scattering effect, any mechanical (or thermal) change in the fiber material, e.g., caused by compression, tension, shear, or any deformation changes, can be recorded because the Brillouin frequency will change. This enables one to detect any mechanical change along an embedded or applied fiber with a certain uncertainty in its position (spatial resolution, see Guideline for Use of Fiber Optic Sensors 2009).

Measurement systems for all three basic optical backscattering effects are commercially available. Depending on the system, sensing areas can be located several

kilometers, in case of Brillouin systems several tens of kilometers, away from the instrument. Because Raman-based systems for temperature measurement are able to bridge a distance of about 10 km, for temperature measurement areas in a longer distance away from the device, Brillouin systems have to be used.

In recent years, more and more applications of distributed sensor to measure strain changes at all points along the optical fiber at acoustic frequencies with the aim to detect acoustic events or vibrations – **DAS, distributed acoustic sensing** – over large areas are reported.

Sensing lengths of up to 50 km can be reached. For a 40 km-long fiber, acoustic signals can be measured that vary at frequencies up to 1.25 kHz. Shorter fiber lengths allow higher acquisition rates, e.g., up to 10 kHz. Special technology modifications allow the sensor to act as a continuous microphone, up to 50 km in fiber length, designed to “monitor” signals into the ultrasonic frequency range (Website [AP Sensing](#)). Acoustic sensing uses again Rayleigh backscatter signals which interact with coherent laser pulses coupled into the optical fiber. Acoustic waves comprising the optical fiber create small changes in the waveguide’s refractive index. These changes affect the backscatter characteristics, it is analyzed, and detected events can then be located by using time-domain techniques with resolution of 1 m or less. This young sensing technology is highly accepted to observe events in the oil and gas industry, for example, in-well applications to detect fracturing processes, flowing and non-flowing wells, flow control of pipelines, leak detection, distributed seismic detection, and/or any movements in security areas (Hussels et al. 2016; Johannessen et al. 2012; Michlmayr et al. 2017). Devices that record distributed acoustic fields at every location along a single-mode or multimode optical fiber, called iDAS ([Silixa](#)), are available on the market. A frequency range of less than 1 mHz to over 100 kHz and a spatial resolution down to 1 m are achievable (Website [iDAS](#)).

Some research is being done to develop distributed humidity sensors with sensitive areas along the optical fiber; however the available power to detect humidity changes along the whole fiber length limits the total sensing lengths to a couple of hundred meters (Liehr et al. 2017).

All the above described sensing solutions can be used to evaluate the operational function and/or the integrity of structures or areas that have to be observed or show certain risk. In the following, only few highlights are mentioned.

---

## Application Examples

### Sensors for Local and Quasi-distributed Measurements

#### Use of FBG Sensors for Static and Low-Frequency Strain Changes

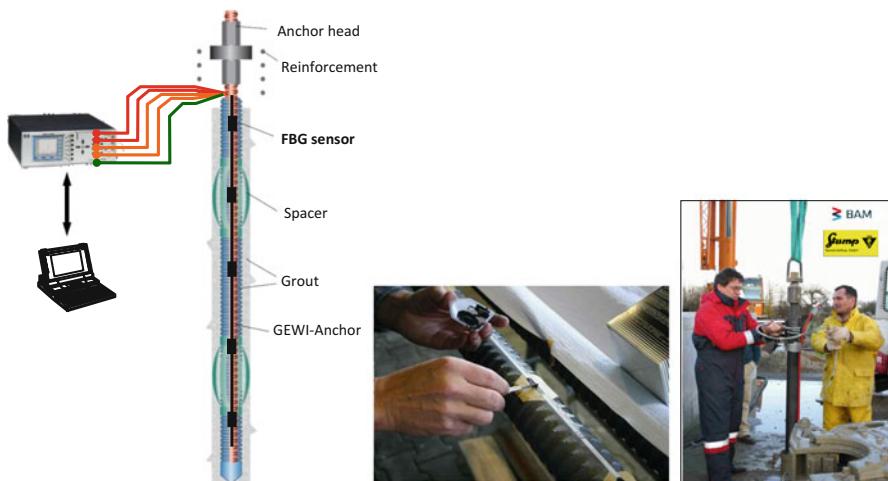
##### Evaluation of Bearing Behavior of Steel Anchors

To evaluate the nonuniform deformation behavior of extended structures or structure components, usually a number of resistance strain gages are applied to those zones in the structure where significant deformation is expected. This includes routing of

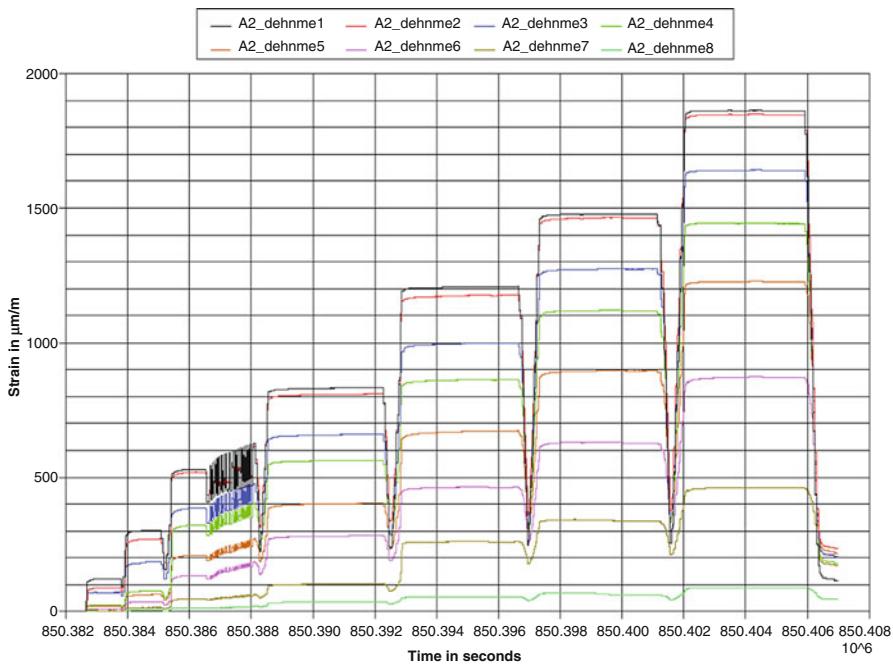
many electrical cables which need some space. If there is no space or if cabling is not acceptable, use of a chain of fiber Bragg gratings offers decisive advantages. Using the example of evaluation of the bearing behavior of steel anchors fixed in the ground, the efficiency of FBG strain sensors will be demonstrated.

The classical way to evaluate the bonding behavior of the fixed anchor length is to measure the resulting forces at the head of the strained anchor. This test method delivers only integral information whether the introduced anchor forces will be transferred into the soil area or not. However, it does not provide any information how the skin friction is distributed along the steel anchor and finally to what amount is the length of the anchor involved in the load bearing. In order to get information on the force distribution along the anchors, the strain distribution along the anchor can be measured by fiber Bragg grating sensors (FBG arrays) attached to the surface of steel anchor. As explained in Fig. 5, strain or temperature change in the FBG leads to a shift of the peak wavelength  $\lambda_B$ . This method was used to evaluate the strain distribution in micro piles (type GEWI® 63.5 mm Z-32.1-9/1), which were used for fixing of foundation plates of a sluice in the German river Weser against uplift (Habel and Krebber 2011). Figure 17 shows the application of a FBG sensor to the steel anchor in the anchor factory. The gage length of each sensor in the chain was 200 mm; the distance between the sensors was between 750 mm and 1,500 mm. Each bar was equipped with two sensor arrays for redundancy reasons. The optical fiber containing all FBG sensors is connected with the leading cable and finally connected to the interrogating device.

After installation of the anchor and complete curing of the grout, pullout tests were carried out. The strain distribution along the fixed anchor length was measured during tensile test. Figure 18 shows the strain development in the anchor at the



**Fig. 17** Measurement concept to evaluate the anchor deformation (left), application of FBG sensor to the anchor steel (middle), and installation of the anchor in the borehole (right). (Source: Sheet of SUSPA GmbH Germany, graphic left: D. Hofmann, BAM Berlin, Germany)



**Fig. 18** Measured strain distribution in the anchor during test loading cycles. The curves show the strain in the attached FBG sensors; the strain levels correspond to the load history over time

defined measurement points where the FBG sensors were attached and for different load steps.

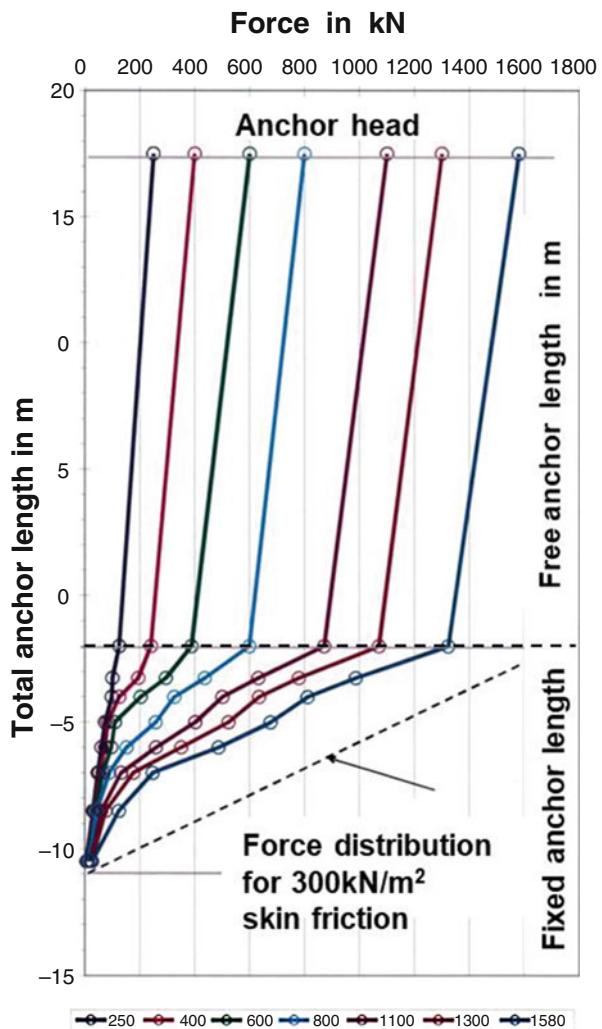
From strain measured in the fiber Bragg grating sensors, the force distribution along the anchor can be calculated. Figure 19 shows the linear force distribution along the non-fixed anchor rod (free anchor length) above the dashed line. In the bottom part (below the dashed line), the force distribution in the fixed anchor area is shown, which represents the load transfer into the soil (load bearing). It is obvious that no evaluation of the integrity of anchor bonding in soil is possible without getting the strain distribution along the steel anchor.

This example concerned a nondestructive evaluation on the basis of rather static strain measurements, although the evaluation of dynamic strain changes is also possible (see the change from the third to the fourth load level in Fig. 18). Another example is to show how the dynamic capability of FBG sensors can be exploited.

### FBG Sensors for Strain-Based Dynamic Responses

Damage of laminated composite material systems or any lack of their integrity may lead to dramatic decrease of the structure's lifetime. Damage mechanisms begin with matrix cracking and fiber breakage, progressing to significant matrix deformations and extensive delamination. Occurring failure modes during damage progress and their interactions enable prediction and evaluation of damage states within a laminate. Early

**Fig. 19** Force distribution in the fixed anchor due to the load transfer to soil (Habel and Krebber 2011)



detection of arising damage mechanisms or evaluation of the severity of post-impact integrity changes by integrated sensors is highly desirable and therefore the object of intense research. Extremely thin optical fibers with inscribed FBG sensors embedded into laminates are of great benefit due to their proximity to the damage features with, at the same time, a minimum of perturbation to the surrounding material microstructure.

FBG sensors are also able to measure dynamic strain by measuring the shift of the Bragg wavelength  $\lambda_B$ . However, evaluation of the spectral response of the FBG delivers much greater depth of information on the strain distribution at the location of the sensor. The following example demonstrates a highly sophisticated methodology to evaluate subsurface (non-visible) damage and possible post-impact residual strain changes by embedded FBG sensors.

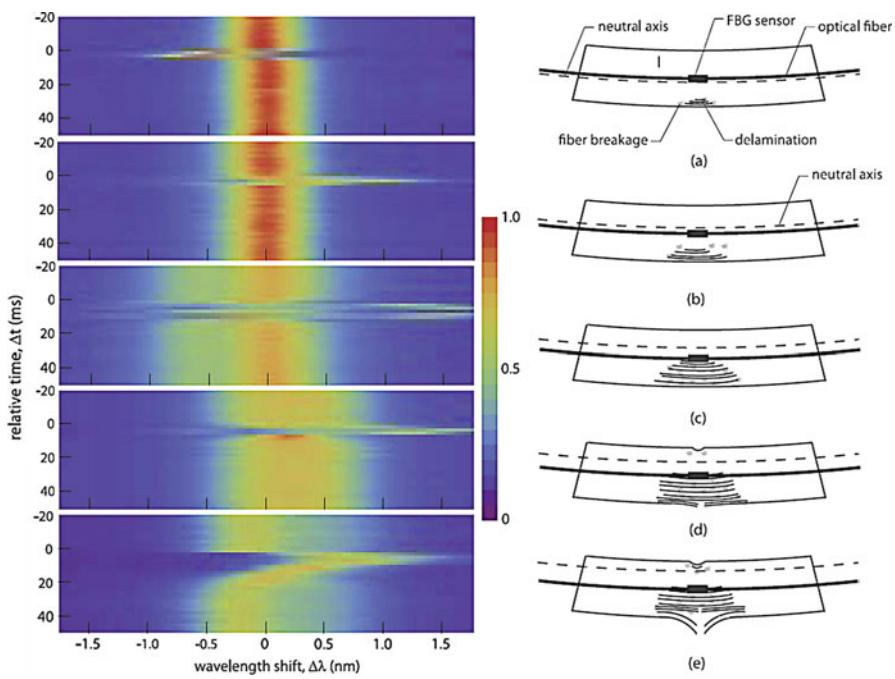
Full-spectral interrogation of a FBG sensor up to 926 Hz was performed while the sensor was embedded in a twill weave carbon fiber-epoxy composite laminate. The polyimide-coated optical fiber was embedded at the laminate midplane during layup of the laminate, and the resulting specimen thickness was 4 mm (Propst et al. 2010). In case of low-velocity impacts to the laminate, no useable results from peak wavelength data were achieved; however, the dynamic full-spectral data response could be correlated to the progression of damage in the laminate.

Damage to the specimen was applied by multiple low-velocity impact events by use of a drop-tower impactor, consisting of a 19 mm diameter hemispherical steel impacting probe on a 5.5 kg aluminum crosshead. Full-spectrum measurements at high scan rates required a specialized system to interrogate the salient features of the Bragg spectrum. It can be quite complicated and different for each sensor due to imponderable influences from the embedding process. The interrogator (readout unit) tracks changes of the spectrum in response to impact and other potential damage-inducing events. The system must work very fast, store and display very large data sets, and must allow high-speed communications.

Changes in the damage status are calculated from cross-correlations between measured spectra during impact and the initial (“undeformed”) spectrum for that same FBG. Results of full-spectral interrogation at 534 Hz and evaluation in a time window of 70 ms around the impact event are shown in Fig. 20. The data sets are visualized by two-dimensional false-color intensity arrays constructed for single strikes and capture the transient response of the FBG during and immediately following the impact event. Applying cross-correlation of the spectra with the initial spectrum, overall trends in the axial tensile, and compressive strains can be observed quite clearly. These images represent plots of the axial strain in the FBG through the wavelength shifts (with color indicating the amount of transverse compression on the sensor). These spectra can be considered as typical responses measured throughout the lifetime of the specimen.

The full-spectral interrogation of the equilibrium, residual strain state permitted the classification of the laminate lifetime into five regimes and the interpretation of the laminate damage state within each regime as shown in Fig. 20, right. It is possible to identify different damage states from the FBG sensor full-spectral response without the need for information on the loading conditions and global response of the laminate. This methodology can be applied to multiple sensors within one fiber and a single laminate. Multiplexed sensors would permit the separation of global and local effects on the optical fiber, which means, the identification of global response mechanisms to vibration of the laminate including a better resolved damage region after a given strike. If embedded between different plies, FBG sensors can observe the progression of delamination through the thickness of the laminate.

In order to exploit the full-spectral capabilities of the interrogation system, research is focused on increasing the data acquisition rate to obtain better resolution of the transient spectral profiles immediately after an impact and to make more precise estimates of maximum strain values during the early impact events. Recent sophisticated investigations based on a tunable MEMS photodiode enabled the interrogation of up to 13 FBG sensors in a single measurement (van Hoe et al.



**Fig. 20** Left: Cross-correlation of the initial FBG spectrum with spectra obtained from different strikes. Right: Estimated damage states during different impact regimes (a) through (e) (bending of the plate occurs in two perpendicular directions)

2017). This system is capable of reading out FBG spectral data sets up to 100 kHz and allows high accuracy both in terms of spectral and temporal resolution. Better visualization of damage zones and progress of damage states is possible.

### Acoustic Emission Sensing in High-Voltage Facilities Using Interferometric Sensor Concept

Partial discharges (PD) generate extremely small material oscillations from which mechanical waves propagate through the insulation material. Such acoustic events are usually detected by dynamically sensitive sensors, e.g., piezoelectric transducers which are placed at earth potential (outside of the insulation material) and convert the acoustic energy into an electrical signal (see ▶ Chap. 4, “Acoustic Emission”). Such acoustic detection is widely used in high-voltage (HV) transformers or gas-insulated substations; however, it requires various electronic components and electric cables that attenuate the electronic signals and makes their use less practical in most cases. To avoid such limitations, optical evaluation methods based on optical waveguides are of increasing importance. Optical fiber sensors can be placed much closer to potential PD sources. They provide all-optical signal transduction and transmission, inherent electromagnetic immunity, and the ability to operate in strong electric fields. The example described here shows the ability to detect and evaluate the intensity of

acoustic waves inside of elastomeric high-voltage insulations. Figure 21 left shows an optical fiber coil made of a bend-insensitive sensor fiber adapted to the electrical insulation material. The sensor is dielectric and small enough to integrate it into the insulation. The optical fiber is tightly wound around a cylindrical core made of silicon rubber to form a Sagnac interferometer. If acoustic emissions occur due to beginning partial discharges (PD), ultrasonic waves cause dynamic changes in the optical path length of the sensor (Rohwetter and Habel 2013; Rohwetter et al. 2015). The detection limit is low enough to detect onset of PD and evaluation of development of such disturbances which allow prevention of catastrophic failure of high-voltage facilities.

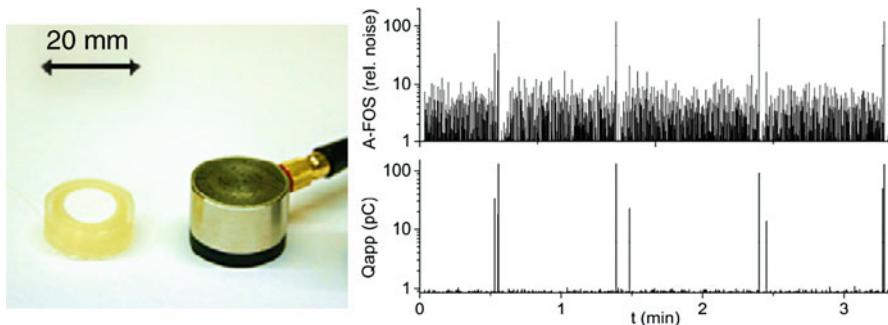
A very important advantage of this sensor technology is that charge carrier activities in dielectrics can be investigated and evaluated. Electrical methods for PD detection based on capacitive coupling method, ultrahigh frequency (UHF) method, or electromagnetic wave coupling method cannot efficiently be used to investigate electrical insulations under direct current (DC) stress. The increased use of DC high-voltage cable lines requires extensive evaluation of the role of the generation of DC acoustic emission. Fiber sensors allow these investigations in a unique way.

Another application to detect failure in expensive high-voltage facilities is to monitor switchgears. Integrated diagnostic hardware based on optical fiber sensors allows evaluation of the sound characteristics of switchgears and following quantification of damaging potential (Malik et al. 2017).

### **FBG Sensor Chains for Evaluation of Temperature Distribution in High-Voltage Facilities**

The advantage of tiny optical fiber sensors, which can be attached to or embedded into components at high-voltage potential, such as windings of high-power transformers or electrical machines, can deliver valuable information about their operation quality. An important measurand is the temperature.

Power generators are operated below the limit temperature with respect to their insulation class specification. It is therefore essential to know the temperature inside



**Fig. 21** Left picture: Photographs of sensors to detect acoustic events: Acousto-optical fiber sensor A-FOS (left) and the piezoelectric sensor (Vallen VS150-M, right); Right picture: Measured signals under direct current (DC) stress taken from A-FOS (above) and from piezoelectric sensor (below)

the machine. Measurement and evaluation of hot spot temperature development at the stator coils enables overheating control and can be exploited to optimize the operation performance. FBG temperature sensors, as single sensor or in a chain (FBG array), allow evaluation of the operation temperature of the high-voltage winding coils. Figure 22 shows two stator bars of an air-cooled power generator with integrated FBG temperature sensors inserted into the slots of the stator core. This online temperature monitoring of the stator winding enables one to evaluate and prove the integrity of the stator winding under load nondestructively.

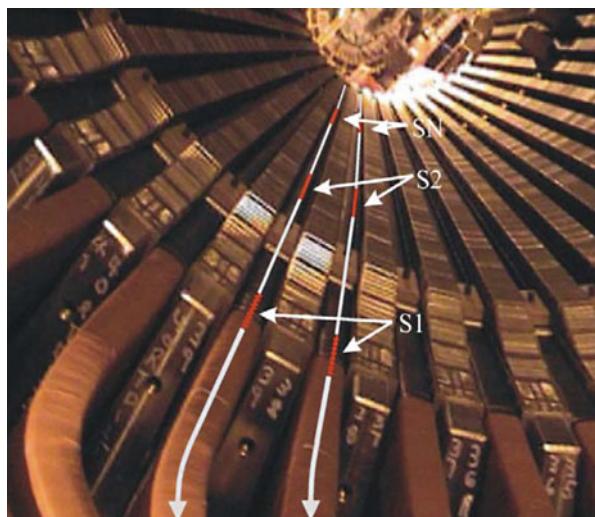
FBG strain sensors can also be used to observe the dynamic behavior of rotating HV machines (generators or motors) during test and during operation. Another objective of the dynamic FBG sensors' use is to evaluate the mechanical behavior, mainly the sound behavior of switchgears during the switching process. A change of the sound characteristic can be understood as a warning signal that the switching element is defective or damaged in one component of the drive mechanism. Such integrated optical fiber sensors provide customers information on the health of the most critical parts of the generator insulation. More diagnostic methods are described in Czichos (2013).

Among the application of FBG sensors for strain, temperature, and vibration measurement, they can also be designed to detect (or even measure) humidity changes and corrosion progress in steel-reinforced concrete components. An embedded FBG sensor is subject to change its Bragg wavelength, when corrosion occurs because corrosion products lead to increasing pressure in the bonding zone between rebars and concrete matrix (Tan et al. 2016).

### Fabry-Perot Sensors for Investigation of High-Performance Materials

If the microstructure of homogeneous materials or complex composite materials have to be investigated, very small sensors are necessary that can be embedded. Additionally, if curing processes have to be observed, not only very small but also

**Fig. 22** Installation of FBG temperature sensors in a stator of a power generator prepared for a shop test (S1 through SN shows the position of the temperature sensors) (Habel and Heidmann 2013)

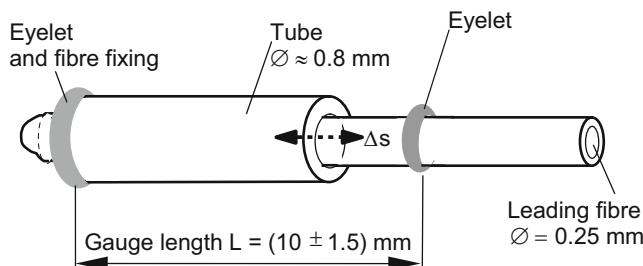


compliant deformation sensors are needed to follow microdeformation processes during development of the solid structure of curing materials.

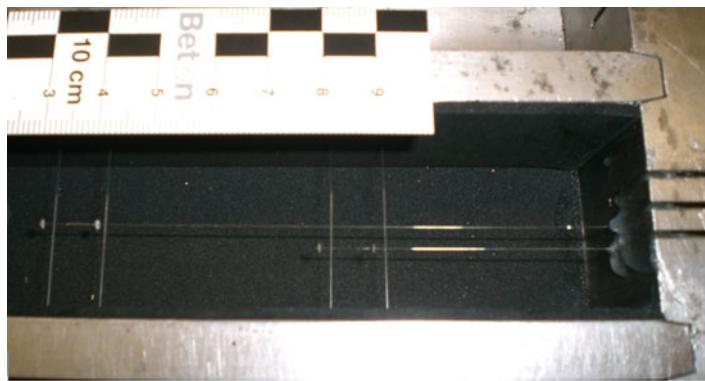
Embedded extrinsic fiber Fabry-Perot sensors (see section “[Sensors for Local Measurements](#)”) offer the ability to measure deformation without significantly influencing the measurement zone. They can be designed as compliant sensors and can follow even the slightest deformations of the material due to their high strain resolution. The following example shows the impressive features of tiny extrinsic fiber Fabry-Perot sensors (EFPI) to evaluate the microdeformation behavior in cementitious matrices.

The motivation for use of EFPI sensors was to evaluate the shrinkage and/or swelling behavior of special mixtures of cement-based building materials at very early ages. The problem is that durability and strength of high-performance concrete is often reduced owing to microcracks occurring as a result of restrained shrinkage and temperature deformations at early ages. In order to get information from the inside of materials, the design of EFPI, shown in Fig. 10, was slightly modified. Figure 23 shows that only one fiber (the reflecting fiber) is positioned inside the capillary; the leading fiber is freely movable.

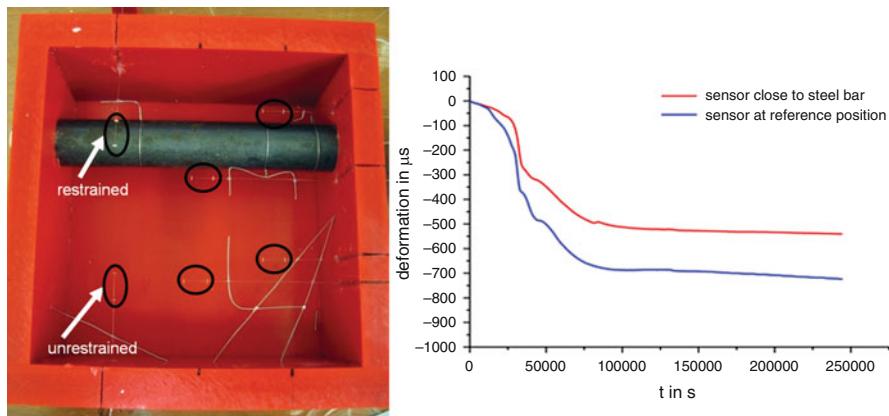
The dimensions of such embeddable sensors are 10 mm in length and outer diameter 0.35 mm. The capillary is usually made from pure fused silica with good resistance to alkaline attacks or from tiny stainless steel tubes used for medical syringes. The leading (movable) fiber links the sensor to the recording instrument. In order to achieve reliable bonding between sensor and material as soon as it sets, small ceramic eyelets (e.g., inner diameter 0.3 mm, outer diameter 2 mm, thickness 0.25 mm) are fixed at the end of the tube and at the leading fiber shortly before it enters the tube. These fixing points are necessary to define the gage length of the sensor and to derive strain values from deformation. The strain limits of the sensor are in the range from  $-2,500 \mu\text{m/m}$  (shrinkage) to  $+5,000 \mu\text{m/m}$  (extension). The strain resolution is about  $0.1 \mu\text{m/m}$ . This design enables one to measure deformation immediately after finishing the casting of the material. Figure 24 shows such EFPI sensors before casting placed to evaluate the microstructure behavior of mortar or concrete from the beginning of the hydration reaction in laboratory.



**Fig. 23** Movable version of an extrinsic Fabry-Perot sensor for deformation measurement without stiffening the object to be measured



**Fig. 24** Left: EFPI sensors prepared for embedment into cementitious material, right: two tubular EFPI sensors with eyelets placed in a standard mold before casting. Such samples are used in laboratories to monitor and optimize high-performance cementitious materials



**Fig. 25** Tubular EFPI sensors with eyelets placed in a silicon mold before casting of cement mortar. Sensors in the bottom area observe and evaluate the curing behavior of unperturbed material; sensors in the upper area measure deformation of mortar influenced by “perturbation”

Such EFPI sensors also allow assessment of restraining effects of reinforcement steel, aggregates, or other inclusions. Figure 25, left, shows sensors close to the rebar (distance, 5 mm) to evaluate the curing process close to such a “perturbation” zone. For comparison purposes, sensors were also placed in the same specimen at such positions where no restraining effects are expected.

Figure 25, right, shows the measurement results during curing of the cementitious mortar. It can clearly be seen that the steel bar has a restraining influence on the deformation (upper curve), particularly during the period of high deformation rate (Schuler et al. 2009; Schuler 2010). This methodology allows systematic investigations and nondestructive evaluation of materials during chemical curing depending on the choice of different steel bars, ribbed and plain ones, with various diameters. It

not only reveals variation of the steel-to-concrete bonding but enables optimization of materials within the required specifications, too.

It is particularly difficult to validate measurement results derived from embedded sensors. One opportunity is to measure deformations simultaneously by a contactless measuring laser vibrometer. The measuring point for the laser beam was a tiny steel plate placed at the front end of the specimen. The steel plate was anchored to the cement paste, and the deformation during shrinkage of the cement paste was recorded. However, it cannot be ensured that this trial represents the real micro-deformation in the matrix material. A better approach is the use of digital radiography. In order to estimate the reliability and repeatability of the results, measurements during the first few hours in particular have been directly validated by using computed x-ray tomographic technology (see ► Chap. 31, “X-Ray Tomography”) (Schuler et al. 2008). All change in the inner part of the matrix material could be made visible and compared with the data recorded by the EFPI sensors.

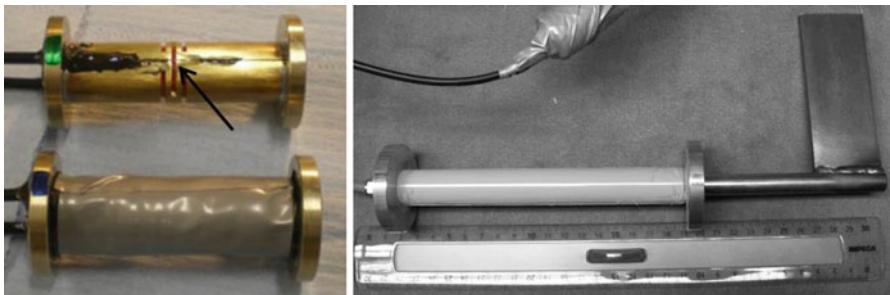
### Fabry-Perot Sensors for Evaluation of Integrity of Large Concrete Piles

The use of EFPI sensors described above for monitoring in laboratories requires certain expertise to embed the filigree sensors without damaging them. EFPI sensors have also excellent dynamic properties so that dynamic or acoustic signals can easily be recorded. If such EFPI sensors are to be used in real concrete components, they should be attached to sensor bodies. The next example shows the use of robust fiber-optic EFPI sensors for evaluation of the long-term behavior of large concrete foundation piles.

Reinforced concrete piles are used to transfer the loads into deeper strata with sufficient bearing capacity if buildings have to be constructed on soft ground. In order to determine the pile's bearing behavior and detect damage, static and dynamic pile tests are usually carried out. The usual method to take dynamic measurements from the pile head to evaluate the bearing capacity and structural integrity can be difficult when very large concrete piles are used. In order to receive more precise information about the pile characteristics, a string of fiber-optic strain wave sensors can be embedded at different levels of the pile. These sensors record the acoustic wave response over the whole pile length caused by an impact at the pile head.

EFPI sensors provide sufficient dynamic properties up to the range of several hundred kilohertz to collect weak strain waves in concrete caused by the static load test and the dilatational wave during dynamic loading. To provide a properly functioning sensor system on the basis of rather fragile optical EFPI elements for application on-site, a robust design which allows installation in harsh environments is necessary.

Figure 26 shows one possible technical solution. The sensing EFPI element is attached to a special flexible sensor body (left picture) and then covered with a protective shell (right picture). The design of the sensor was matched to the concrete mixture (aggregate size) and to the dynamic loading conditions of the pile dimensions. Because bending of reinforcing steel elements could occur, the sensor cage was separated from the reinforcement cage of the pile. For this reason, the already



**Fig. 26** EFPI impact sensor design for integration into reinforced concrete piles. Left: EFPI sensor is attached over the flexure to sense weak acoustic waves inside the concrete pile, right: weldable sensor rod with a protective shell



**Fig. 27** EFPI impact sensors mounted in the cage of reinforcement of a driven pile (left) and a cast-in situ pile (right)

mechanically decoupled steel part of the sensor was suited as mounting. A steel plate was therefore welded at the end of the sensor rod, and this plate was welded to the reinforcement elements.

Figure 27 shows sensors integrated into a cage of a precast driven pile (left picture) and into a cast-in situ pile (right picture). It is worth noting that all embedded sensor rods did withstand the conditions during the driving process and during concreting the cast-in situ pile. More details can be found in Schilder et al. (2012).

Integration of these optical fiber sensors allows highly sensitive detection of strain waves in large concrete structures which enables nondestructive evaluation of the integrity of load-bearing components. In the case of driven concrete piles, the whole

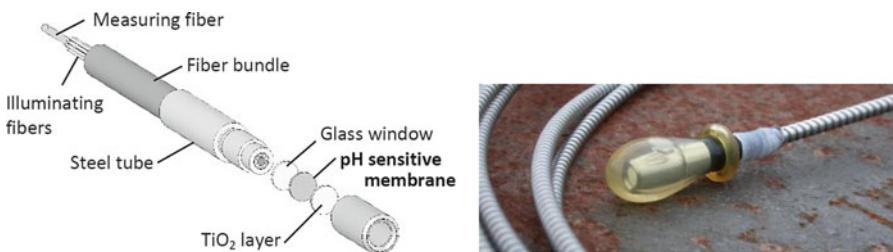
driving process can be supervised, and possible damage can be immediately detected. After completion of pile construction, low-strain and/or high-strain dynamic tests deliver reliable data about the entire pile integrity. These sensors also provide precise information about the deformation profile of such piles during static loading which allows evaluation of the load-bearing capacity especially in difficult soils. And finally, such sensor-equipped concrete structures deliver valuable indication on how to optimize the design of future load-bearing structure components.

### **Fiber-Optic Sensor Probe to Detect Potential Danger of Corrosion in Steel-Reinforced Concrete Structures**

Corrosion in steel-reinforced concrete structures is one of the most relevant damaging processes. Corrosion appears if the pH value of the concrete matrix decreases due to deterioration of the calcium hydroxide layer on the steel surface. The safest information about potential danger of steel corrosion is the knowledge of the pH value of concrete. The lifetime of steel-reinforced concrete structures depends strongly on their pH state as embedded steels in concrete structures are only passive at pH values higher than 9. For this reason, long-term monitoring of pH values in the range of 9–13 with a resolution of about 0.5 pH units is relevant for early detection of potential corrosion condition. Commercially available structure-integrated sensors for nondestructive evaluation of steel corrosion state in concrete structures do not always sufficiently match the in situ requirements. Because the pH value as a chemical parameter is of major importance for the health of steel-reinforced and prestressed concrete structures, it is recommended to observe directly the development of the pH value.

There is the possibility to detect chemical species and chemical status of materials by using optical spectroscopy. Fiber-optic-based sensors are a promising technology for corrosion monitoring because they offer a large number of attractive features such as small size, flexibility, geometric versatility, resistance in corrosive and hazardous environments, no signal interference due to present moisture, in situ and nondestructive measurement, and also immunity against lightning strikes. Depending on the pH value, an indicator immobilized in a solid substrate and placed in the optical path changes the color (wavelength, hence the spectral response). Intense investigations revealed that the absorption method provides reliable measurement results, because instability problems resulting from decrease of the indicator concentration due to photodegradation or leaching out, drifts of the light source intensity, or bending of optical fibers had to be managed. A pH change leads to a wavelength change in the pH-sensitive substrate; two different wavelengths are detected to exclude aging effects (ratiometric method). The membrane is sensitive in the pH range from 9.5 to 12.0; pH values can be resolved with 0.1–0.2 pH level, whereas the highest resolution is achieved in the middle of the measurement range (between pH 9.7 and 11.0).

The sensor probe (see Fig. 28) can be embedded into the concrete matrix and observe changes in pH over years. Several sensor probes can be staggered to detect the progression of pH decrease.



**Fig. 28** Left: Exploded view of the pH sensor head (8 mm in diameter); the head is pluggable and can be replaced. Right: pH sensor prepared for embedment in concrete; the pH-sensitive membrane is protected by a small watertight topcoat to avoid drying out

The measurement principle and some more details are described in Dantan and Habel (2006) and Habel et al. (2011). There are some practical requirements. The most challenging requirement concerns the long-term stability of the membrane under strong alkaline conditions within the pH range of 13 and 9 over a period of at least 25 years. The sensor has to be integrated in harsh environments and inaccessible places. It must be guaranteed that the protected sensitive membrane has sufficient contact with the concrete matrix to pick up pH changes. One particular condition is that the pH-sensitive membrane must not dry out. This is managed by protecting the sensor head with a small watertight topcoat; after installation this requirement is mostly fulfilled in hydraulic engineering and geotechnical applications.

pH sensors can be attached to structure elements before the poring of concrete, for example, to anchors before installation and grouting (Fig. 29, left). Another option is the embedment of pH sensors in prefabricated cubes (Fig. 28, right). The different depths of the pH-sensitive membranes enable evaluation of the progress of the carbonation front and estimation of the possible onset of corrosion.

Such prefabricated sensor-equipped concrete cubes can be produced on-site using the real concrete material and then immediately after setting be installed into the structure to be observed. This was done in a newly erected cooling tower, where the shell was not protected against acidic condensate. To evaluate the durability, embedded pH sensors help to avoid damage in reinforced concrete components.

### Segmented Optical Fiber Sensors for Evaluation of Bonding of Large Steel Anchors

Evaluation of anchor rods by using FBG sensor arrays to reveal the strain/force distribution along fixed anchors was shown in this section above. If the bonding quality and carrying capacity of heavily loaded anchors is to be checked over large bonding areas, an alternative quasi-distributed sensing technique is used: the methodology of quasi-distributed sensing based on segmented optical fiber sensors (see Fig. 13). Optical fibers containing prefabricated measurement segments are embedded into the structure component to be evaluated. The lengths of such segments can be defined according to the measurement task and can be between a few centimeters up to several meters. The measurement principle is explained above (see Fig. 13); it is based on the time-of-flight measurement of a launched optical pulse and its

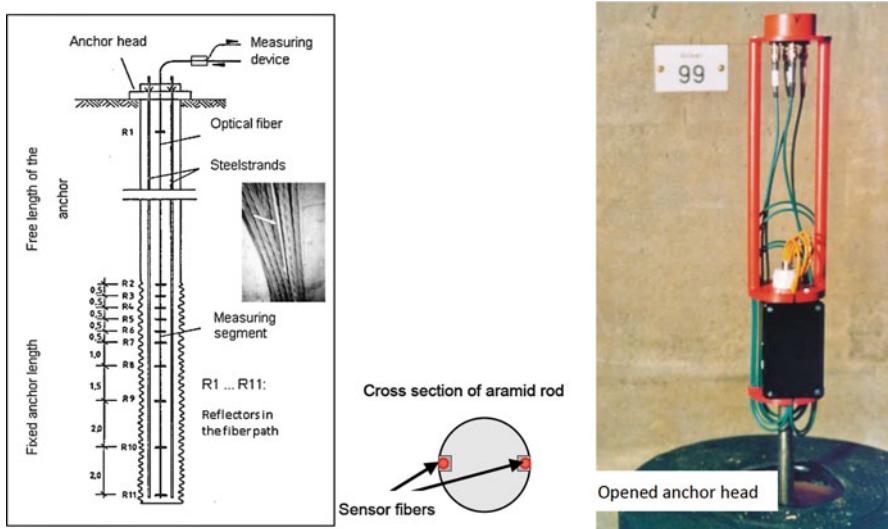


**Fig. 29** Prefabricated anchors with fixed pH sensors (left picture, arrows). Right: Concrete cube with three embedded fiber-optic pH sensors. The concrete coverage (from the bottom of the cube) is 5 mm, 10 mm, and 15 mm

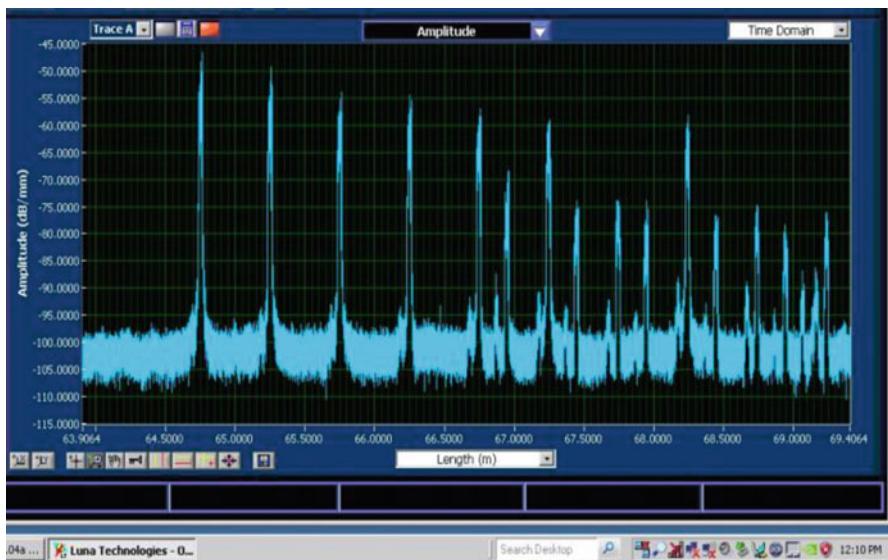
corresponding answer back-reflected from a mirror in the fiber. There is a linear relationship between the length change of a segment  $\Delta l$  and the change in the travel time  $\Delta t$  of the pulse:  $\Delta t = K \odot \Delta l$ , where  $K$  is a proportionality factor considering material characteristics and environmental influences, such as the index of refraction of the optical fiber, elasto-optical material effect, and temperature. If the recorded time of flight increases, the reflecting point is further away from the source of light which means extension of the measurement segment. If the travel time is shorter, the observed component shranked or was compressed. Figure 30, left, shows the sensor arrangement integrated into the complete length of heavy steel strand anchors which are under a permanent tensile load of 4,500 kN (Dietz et al. 2001). Two optical fibers containing 12 measurement segments are embedded in an aramid rod which is located in the center of an anchor formed by 36 steel strands (insert picture). Eleven sensor segments are directly bonded to the grout in the fixed anchor length area. The location of the sensor fibers in the aramid rod is shown in Fig. 30, middle.

Figure 31 shows sensor signals back-reflected from measurement points along one rock anchor of a gravity dam. Based on the anchor length of 70 m, deformations at the measurement points of 20  $\mu\text{m}$  can be detected. This is sufficient to recognize any change in the bonding integrity or changes in the ground situation, e.g., development of cracks or cleaving zones in the rock area.

Because such heavy anchors are not common and the tendon bond behavior over a period of several years was not known, integrated sensor systems were necessary to get permission for this single case to instrument these anchors. As learnt from measurements of the deformation behavior of all parts of the steel anchor over many years, this methodology is the only chance to evaluate the real bonding behavior in the ground and/or rock. Observation of the tensile force situation does not deliver correct information about the bonding integrity because temperature changes of the gravity dam can lead to expansion of the free anchor length and, as a consequence, to misinterpretation of the bonding behavior.



**Fig. 30** Left: Schematic description of sensor equipment heavy steel anchors with segmented optical fiber sensors defined by the place of reflectors; arrow in the insert pictures shows the centered aramid sensor rod. Middle: aramid rod with embedded optical fibers, right: Connection point at the sensor head to perform inspection measurements



**Fig. 31** Evaluation of force distribution in 4,500 kN rock anchors installed in a gravity dam via strain measurement along the steel anchor. The peaks represent the position of measurement points in the fixed anchor area. Horizontal shift of the peaks shows deformation of the anchor steel from which a possible loss of the anchor fixing can be concluded. This method allows one to prove the integrity or the effectiveness of anchor fixation

Details about design of the anchors, manufacturing, and on-site installations are described in Dietz et al. (2001), Habel et al. (1999), and Habel and Gutmann (2005).

## Sensors for Fully Distributed Measurements

The use of optical fibers for distributed measurements (also called fully distributed in contrast to quasi-distributed sensors integrating the measurand over a certain length) is the most interesting approach because they offer the capability of measuring the measurand at thousands of points simultaneously using a simple optical fiber as the sensing element. Distributed sensors are preferably used to measure strain and temperature changes at any position along a single optical fiber with a gage length in the range of a few meters up to several tens of kilometers sensing length.

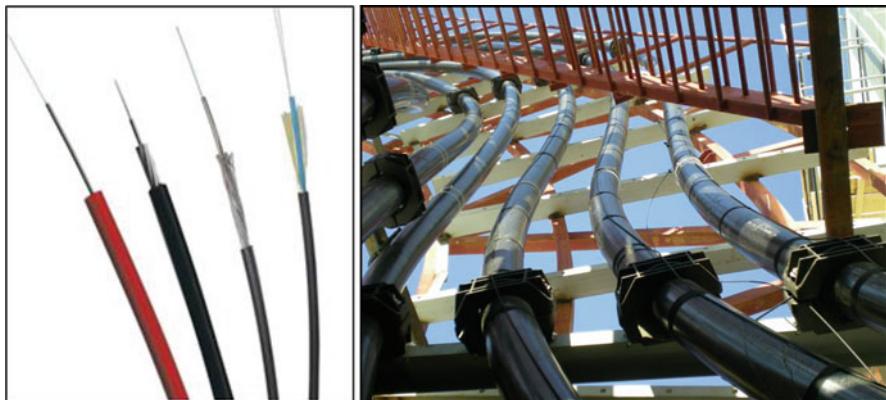
Geotechnical areas are naturally of widespread character, such as slopes, rock areas, and areas of tectonic or mining subsidence. Structures in geotechnical areas, such as tunnels, gas and oil pipelines, and sewer pipes, can be several kilometers long. These structures when used in tectonically difficult areas need continuous automatic monitoring to evaluate proper functioning (see ► Chap. 44, “Structural Health Monitoring”). If these structures, for example, long-span bridges, are designed from composite or high-performance materials, the integrity and their proper characteristics must be continuously evaluated. On the other hand, high-voltage cables, especially submarine high-voltage power cables or links to offshore wind parks, should be monitored to detect hot spots, possible moisture penetrating in cable joints, or damage of the cable sheath by anchors or other mechanical attacks.

### Distributed Temperature Measurement

Besides distributed temperature measurements along pipelines or in tunnels and coal mining to detect early damage or fire, optical fibers are integrated into high-voltage cables to detect hot spots and to control the thermally based stress regime of such cables. There are different concepts to integrate optical fiber sensors (Website [AP Sensing](#); Ukil et al. 2012; Website [NKT Photonics](#)). Figure 32 shows two examples of integrated optical fiber sensors to evaluate the proper functioning of such cables. This sensing technology is quite successful as a diagnostic method because users do not get information about long-term reliability, long-term stability, and/or availability at minimum maintenance in any other way.

### Distributed Strain Measurement

Fiber-optic distributed strain sensing is the most frequently used distributed sensing technology. Such sensor systems collect spatially continuous strain, and temperature data (distributed data) are therefore able to monitor full strain fields and temperature gradients over extended areas. This allows insight into the behaviors of designed structures, for example, how loads are distributed throughout a structure, which legacy technologies cannot accomplish. One of the most outstanding opportunities using this sensing technology is the application of distributed data for model validation and improving design processes besides ensuring the safety of structures.



**Fig. 32** Sensor cable for distributed temperature monitoring (left) integrated into high-power cable. (Photos: LIOS Technology GmbH (Website [NKT Photonics](#)))

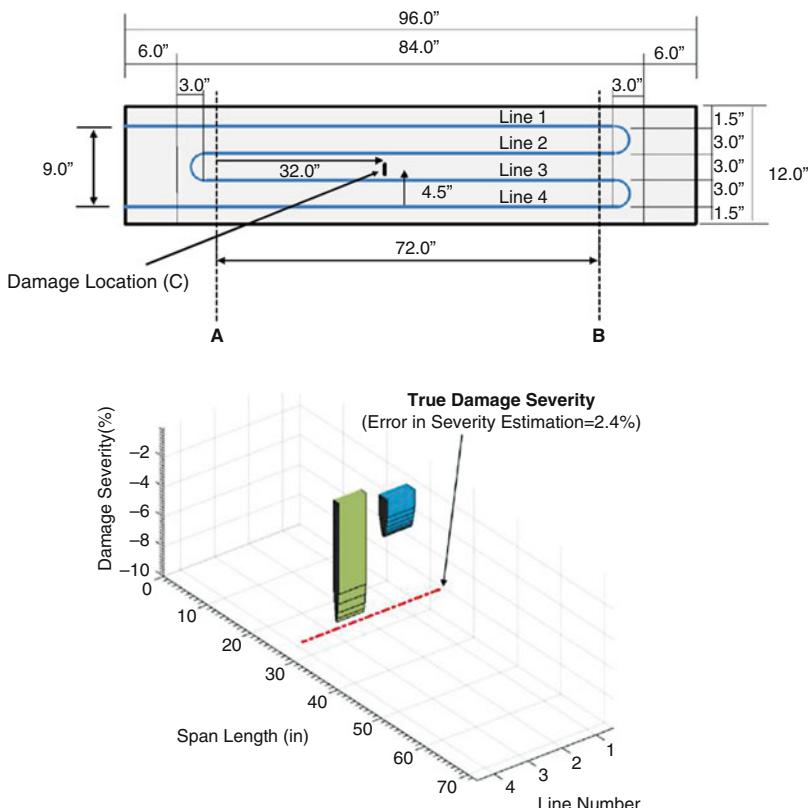
The following example shows how engineers can more efficiently and confidently validate their theoretical assumptions or models: finite element analysis or thermal models of complex structures. Based on such experimental investigations, engineers can refine their models, and a better nondestructive evaluation can be performed.

Complex concrete structures have to be designed, calculated, and validated. It is necessary to gain an understanding of how commercially available high-performance structures behave. A number of tests are therefore carried out to approve the functionality. This is, for example, mandatory for high-performance materials such as rotor blades of wind turbines, airplane wings, or helicopter blades. Especially modern adaptive structures should fully be understood as to how loads are redistributed throughout the components during operation. To find optimized shape conditions for adaptive structures, the strain and load data can be fed into a feedback system, which reacts in real time to achieve maximum efficiency in operation.

This sensing strategy also allows one to know when and where damage occurs during operation. Any damage may be expressed in terms of local decreases in the flexural stiffness of structural members, which cause singularities in the measured strain profiles. Not only to detect damage zones but also determine the location of damage and its severity, optical fibers can be integrated into, for example, composite structures for NDE purposes. Exploiting this technology, their ability to detect weak structure behavior both at critical points and everywhere in between, minimal post-processing is required since strain measurements are directly used to detect damage.

The fibers should be installed in a meandering pattern so that the whole structure is covered (see Fig. 33). Depending on the quality of the readout system, spatial resolution in the millimeter range can be achieved.

Not only mechanical measurands are important to achieve optimum operation, also temperature distribution during production or operation can influence mechanical characteristics of highly loaded materials. This has to be taken in the focus for temperature ranges from cryogenic temperatures up to several hundred degrees Celsius.

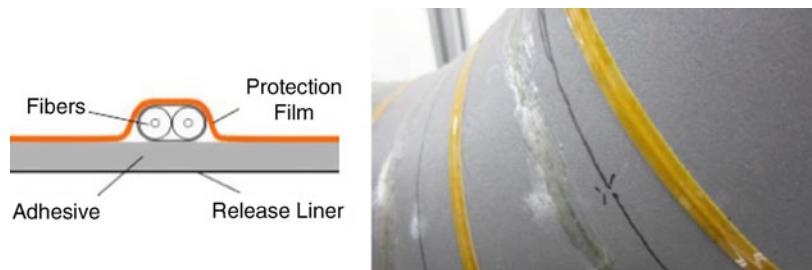


**Fig. 33** Left: Test sample of a composite material indicating the layout of the installed optical fiber and the location of the damage (letter “C”). Right: Existing damage severity derived from the optical fiber-based NDE system (Website [Sensuron](#))

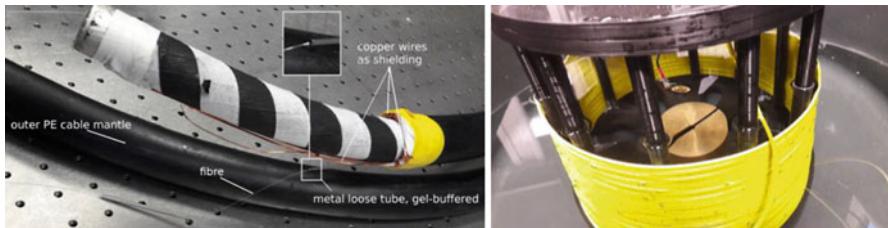
Another important advantage of embedded distributed optical fiber sensors is the manufacturing process, i.e., that the layup quality can be verified. Based on these data, the manufacturing process of composite materials can be improved. The presence of defects during production or operation can be revealed early; common resistance strain gages are not small enough and cannot be embedded and, on the other hand, would not be able to detect a defect unless it happened around the place of the local sensor.

In order to enable easy installation and excellent durability, optical fiber sensing cables are in some way prefabricated to provide an optimum transfer of strain and temperature between the object that is being monitored and the sensing fiber(s) as well as to avoid any damage during installation. A common design is a sticking tape-like sensing component as shown in Fig. 34.

A number of other application examples for distributed sensors in civil and geotechnical engineering are given in Habel and Krebber (2011), Website [Micron Optics](#), and Website [OZ Optics](#).



**Fig. 34** Left: FN-SSL-3 tape sensing fiber consisting of two standard SM fibers on a single tape, designed for temperature and/or strain measurements. Right: Application of a tape sensing fiber spirally adhered to a steel pipe for detection of strain. (Courtesy of Neubrex Co., Ltd., Japan)



**Fig. 35** Power cable specimen with embedded optical fiber placed in a gel-buffered loose metal tube (left); test arrangement to evaluate the influence of acoustic waves on the sensor fiber in a water basin (right)

### Distributed Acoustic Sensing

This example shows an approach to evaluate the proper functioning of submarine high-voltage power cables. Detection of damage due to anchor drops, construction, and trawler shipping would produce acoustic waves or vibration events; its early detection and localization helps to prevent severe damage or failure or even to indicate who is causing the damage. One solution is based on direct detection coherent optical time domain reflectometry (C-OTDR), which measures local variations in Rayleigh backscatter intensity in optical fibers due to dynamic strain variations along the fiber axis (Hicke and Krebber 2017). Figure 35, left, shows the acoustic test arrangement consisting of a power cable with embedded optical fiber for communication purposes. This fiber was placed in a gel-buffered loose metal tube; the metal capillary was then embedded in the outer shielding layer of the cable directly below the polyethylene (PE) cable mantle. Figure 35, right, shows the laboratory arrangement for acoustic tests under water. The sensor fiber was loosely placed in a gel-filled tube, whereas the tube was embedded in a soft outer jacket. A 20 m-long patch cable was loosely wrapped around circularly arranged plastic rods on a circular base. The whole arrangement was then placed in a basin filled with water.

The test rigs were mechanically excited by finger tapping and by dropping a screwdriver into the water basin and removing it by hand. Already with a simple

loose metal tube fiber cable in the outer shielding rigidly coupled to the bulk power cable, distinct low-amplitude acoustic events can be resolved in the frequency domain, i.e., a good sensitivity to external vibrations is possible over a certain cable length.

---

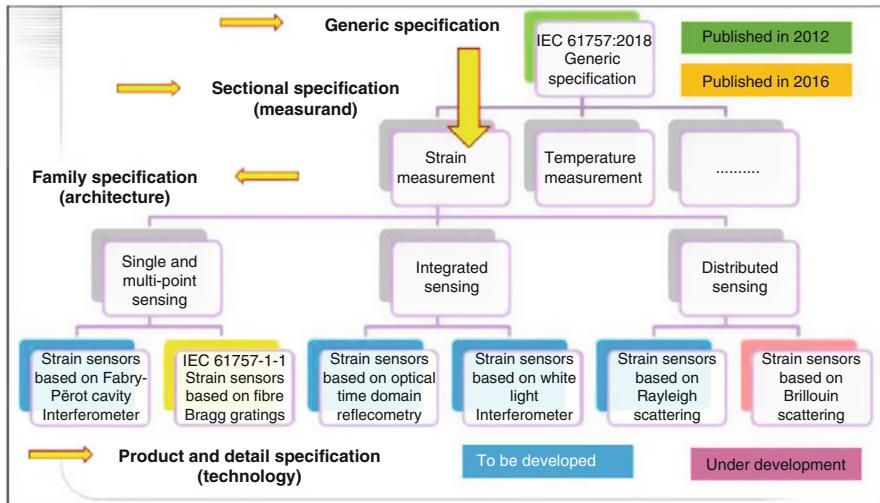
## Standardization

A number of companies providing fiber-optic sensor solutions offer proprietary guidelines or even courses to learn the use of fiber-optic sensors. However, generally there were some barriers for industrial users in the past although the scientific and technological background for optical fiber sensors is well developed. One barrier can be the price for the whole system; another barrier could be the lack of standards or inappropriate specifications from manufacturers of sensor components. Because users need standardized description of such rather new sensor systems, recommendations as to what aspects must be considered for reliable application and operation of sensors, and finally guidelines how to handle the application of sensors under possibly harsh environmental conditions (Habel 2009): Few guidelines for the appropriate specification and use of fiber-optic sensors have already been published. Several international activities have been launched to encourage the development of additional standards for special applications.

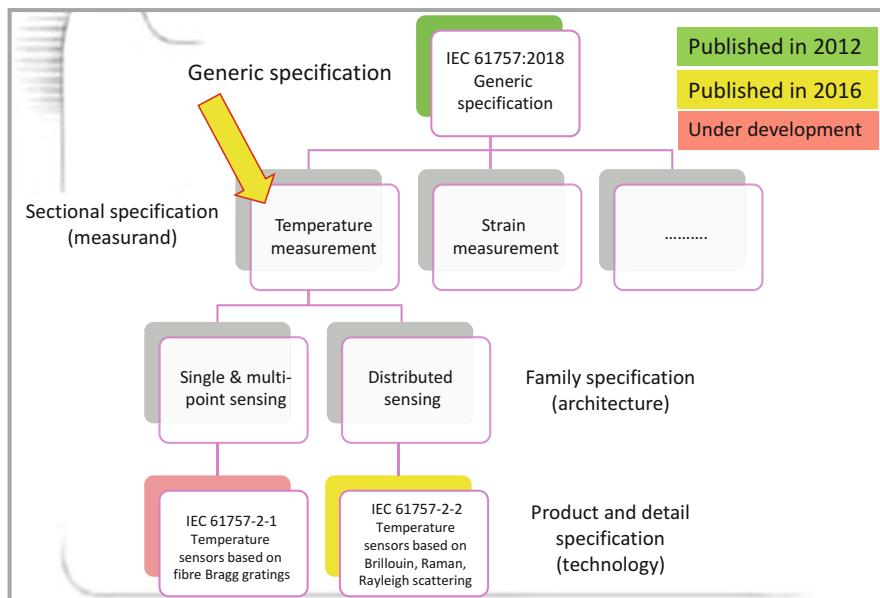
The first standard on generic specification of fiber-optic sensors – IEC 61757 – was published in early 2012; the revised version was published on 25 January 2018 as IEC 61757:2018 Fiber-optic Sensors – Generic Specification ([IEC 61757:2018](#)). This standard helps to use a consistent “standardized” terminology, because not only different users use non-consistent terms and descriptions for the fiber-optic sensor technology, but even all the fiber sensor experts carry the same meaning on what specific terms associated with characterization, validation, and application of sensing systems do convey. This generic standard covers characteristic details related to the respective physical sensor mechanisms, to the sensor response to different measurands, to application, and finally to specific perturbing influences coming from environmental stimulants.

Based on the standard IEC 61757, a fiber-optic standard family was created (see Figs. [36](#) and [37](#)). The next activity in the IEC Subcommittee 86C, Working Group 2, was the development of the IEC 61757-1-1 standard on “Strain measurement – Strain sensors based on fiber Bragg gratings” published in February 2016 ([IEC 61757-1-1](#)). This standard was essentially based on the first bilingual German VDI/VDE 2660 Guideline: “Experimental Stress Analysis – Optical Strain Sensor based on fiber Bragg grating; Basics, Characteristics and its Testing” published in July 2010.

At the same time, the IEC 61757-2-2 standard on “Temperature measurement – Distributed sensing” has been published in May 2016. It is based on the SEAFOM document SEAFOM-MSP-01 “Measurement Specification for Distributed Temperature Sensing (DTS)” originally published in March 2010 by the SEAFOM Measurement Specification Working Group, which is an official liaison partner of IEC SC 86C, WG2 (see Fig. [37](#)).



**Fig. 36** Structure of IEC fiber-optic sensor standards related to strain measurements



**Fig. 37** Structure of fiber-optic sensor standards related to temperature measurements

The next activity in standardization of fiber-optic temperature sensor based on FBG has been done in Germany; the draft of the VDI/VDE Guideline 2660 part 2 on “Optical temperature sensor based on fiber Bragg grating – Fundamentals, characteristics and sensor testing” has been published as a green paper and is being

discussed by European experts. Based on this activity, the corresponding international standard IEC 61757-2-1: “Fiber-optic Sensors-Temperature measurement – Temperature sensors based on fiber Bragg gratings” is under development in the IEC SC 86C, WG 2; a first publication for international experts’ discussion is expected in early 2019. Other activities in the IEC SC 86C, WG 2 are focused on development of IEC standards for distributed strain sensing (IEC 61757-1-2), distributed acoustic sensing (IEC 61757-3-2), and electrical current sensing – polarimetric method (IEC 61757-4-3). The numbering code for fiber-optic sensor standards is shown in Table 1.

There are some more activities driven by other standardization bodies or expert communities, for example, the aeronautic community. Whereas the IEC standards cover rather basic aspects for preparation, testing, validation, and description of sensors and sensor systems, more application-related standards or guidelines with advice to handle fiber-optic sensor systems on-site are developed by industry-related groups. Examples are documents published by Committees of SAE Aerospace, such as ARIR806-5 “Fiber Optic Installation and Maintenance” (2014) or the Aerospace Information Reports AIR6258 “Fiber Optic Sensors for Aerospace Applications” (2015) and AIR6162 “Fusion Splicing for Optical Fibers” (2017) developed by the Aerospace Avionic System Groups Committee AS-3 “Fiber Optics and Applied Photonics.”

It is also worth mentioning the activities in the global optical fiber sensing system (OFSS) task group of the ASTM Technical Committee F36 “Technology and

**Table 1** Numbering code for IEC standards on fiber-optic sensors

IEC 61757 Fiber-optic sensors – generic specification				
Measurand (M)	Technology (T) architecture			
	IEC 61757-M-1 Fiber Bragg grating	IEC 61757-M-2 Distributed sensing	IEC 61757-M-3 Faraday effect	IEC x-M-T
Strain measurement	IEC 61757-1-1 Strain sensors based on fiber Bragg gratings	PNW IEC 61757-1-2 Distributed sensing		
Temperature measurement	CD IEC 61757-2-1 Temperature sensors based on fiber Bragg gratings	IEC 61757-2-2 Distributed sensing		
Acoustic sensing		PNW IEC 61757-3-2 Distributed acoustic sensing		
Current measurement			PNW IEC 61757-4-3 Optical current sensors based on polarimetric method	

CD committee draft, PNW new work item proposal

Underground Utilities,” especially Subcommittee F36.10 “Optical Fiber Systems.” One very helpful document is the ASTM Standard F3092-14 “Standard Terminology Relating to Optical Fiber Sensing Systems” (2014). It includes more than 400 terms commonly used in optical fiber sensing systems. More details about international standardization activities are described in Habel and Jeyapalan (2018). Another summary of international standardization activities can be found in Krohn and Mendez (2017).

---

## Summary and Future Trends

Different optical fiber sensors are used to evaluate the behavior of steel-reinforced and prestressed concrete structures, in high-performance composite structures, and to learn the behavior of new material designs. Not only mechanical quantities are needed to fully understand the behavior under loads and environmental influences, but also the influence of physical and chemical quantities, for example, temperature, humidity, and chemical species, is important to know. More and more optical sensors to measure such physical and chemical quantities, for example, pH value (movable), chloride concentration, and deteriorating chemical processes, are available and offered on the market. However, in contrast to optical sensors for mechanical quantities, optical sensors for physical measurands are rather a niche technology in the market. Even if they have some advantages, the number of intended application is still low and will slowly increase with better knowledge about this technology and corresponding technical equipment and skills to work with optical sensors.

Strong push toward automated, online observation of the structure’s behavior is observed and its importance will grow. Embedded optical fibers, which do not perturb the microstructure around the sensor position and its behavior, will be used more and more. More attention has to be paid to appropriate integration of sensors into the material so that the transfer of the measurand from the material to the sensing part of the optical fiber is optimal. In standards and/or guidelines and also in literature, recommendations on how to approve this load transfer or temperature transfer are given (IEC 61757-1-1; VDI/VDE 2660; Schlüter 2010). Existing testing facilities for optical fiber sensors should be improved, so new testing devices including for mobile use under field conditions can be expected (Santos and Farahi 2015). Expert committees and societies will provide necessary educational and training opportunities in companies, testing institutions, or even in universities to establish this new sensing technology.

Although the basic sensing effects for optical fiber sensors are well developed, new and/or specially adapted materials for optical waveguides provide more variety of application, for example, special optical fibers based on sapphire crystals (sapphire fiber) that can be used for temperature and strain measurements at 1,200 °C and above or specially doped fibers to detect chemical species.

A special class of microstructured fibers, so-called photonic-crystal fibers (PCF) based on the properties of photonic crystals, opens a very wide field in future optical sensing. The light in such PCF is guided by structural modifications, not only by

refractive index differences as used in classical – considered in this chapter – optical fiber sensors. Two specific categories of PCF – the photonic bandgap fiber (confining light by bandgap effects) and the holey fiber (using air holes in their cross sections in different geometrical designs) – have a huge potential to evaluate strain, temperature, bending, chemical, and biological processes. A special type, so-called in-fiber long-period grating (LPG) sensors, can be designed to have exceptional low temperature coefficient while the strain sensitivity can be significantly enhanced. Such sensors show many advantages compared with those in conventional single-mode fibers and index-guiding PCF-based sensors. Because these types are significantly more expensive, the application of this sensing technology will remain a niche technology in NDE of materials and structures (Ju and Jin 2009; Rota-Rodrigo et al. 2013).

Further future developments will focus on improvement of the characteristic values of already established sensors, mainly of FBG-based sensors and all kinds of distributed optical sensors (strain, temperature, and acoustic measurement). Improved digital scan technologies combined with novel optoelectronic architectures of devices will provide better dynamic features, increase of the sensor length to more than 50 km, higher resolution, greater long-term stability, and repeatability of measurements.

More future research should also be focused on finding methodologies to separate temperature influence from mechanical measurands. This is very important for fiber Bragg grating sensors embedded in anisotropic laminates, which should deliver strain information under different thermal conditions. Use of additional mechanically decoupled sensors that only measure temperature to compensate for its influence is not easily achievable without disturbing the laminate structure. Approaches, which use methods of simultaneous determination of temperature and axial strain using the spectrum of a single superstructured FBG, are promising (Schukar et al. 2014).

Reliable long-term evaluation of materials behavior by using embedded or attached sensors is an absolute prerequisite. Any problems with sensors that possibly are delaminated inside the composite or debonded at the surface must be identified. When fiber Bragg grating sensors are used to evaluate structures, observation of the FBG spectrum can reveal problem with the sensing quality. Research to exploit spectral deformation of the reflected FBG spectrum, which is sensitive to partial debonding of the sensor, should be intensified. Developed analytical methods to determine degree and direction of debonding should be applied to qualify and evaluate the interface performance of adhesively applied sensors and to evaluate delamination growth of composite structures (Baitinger et al. 2014). This sophisticated monitoring technology will also help obtain knowledge about the validity of the calibration performed at the time of entry into sensor operation (“self-calibration” check). Embedded and/or applied sensors that are out of reach could tell the user that they do not operate correctly any more.

And finally, future fiber-optic sensor systems will differ from conventional ones, which are based on a set of single sensors linked in a network and provide a flood of information over a long period of measurement. Among a complex net structure of single, quasi-distributed, and distributed sensors, complex signal information will be

more and more filtered, and compressed and specific acquisition algorithms and classification techniques provide individually edited data for quick identification of critical situations. Information about amount and location of the measurand change will be more and more precise. Interaction with other technologies such as integrated optics or other highly integrated sensor technologies – MEMS (micro-electromechanical systems) – will advance rapidly.

Cost development of all components of optical fiber sensors is difficult to estimate. It depends highly on the specification, especially the requirements on characteristic parameter and stability, on specially designed configurations, on distance, and on application conditions. All components developed for the exclusive use of optical fibers as sensors will remain expensive because there is not a mass market for them. However, it must not be forgotten that for a number of optical sensors in one optical fiber (or even thousands in one distributed sensor arrangement), the overall price could be lower than those of conventional systems.

Standardized sensor solutions will find their use in those fields where advantages of optical fiber sensors beat all electrical sensing technologies. Reading standards, not only characterization and specification of an optical fiber sensor system is easy to handle but also the right selection and application.

---

## References

- AP Sensing. Advanced photonics. <https://www.apsensing.com>. Accessed 28 May 2018
- ASTM F3092-14 (2014) Standard terminology relating to optical fiber sensing systems. ASTM International, West Conshohocken
- Baitinger E, Schukar VG, Kusche N, Schilder C (2014) Delamination-diagnosis-method for adhesively surface-applied FBG strain sensors. SPIE 9157:91578O, 4 pp. <https://doi.org/10.1117/12.2059671>
- Czichos H (ed) (2013) Handbook of technical diagnostics fundamentals and application to structures and systems. 1st edn. Springer Berlin Heidelberg. ISBN 978-3-64225849-7
- Dantan D, Habel WR (2006) Monitoring of corrosion protection – a concrete-embeddable pH optode. BFT Int-Concr Plant+Precast Technol (Betonwerk+Fertigteiltechnik) 72:48–55. Springer
- Dietz K, Habel WR, Feddersen I (2001) Eder Dam – stabilisation by permanent rock anchors – monitoring and long term performance. In: Proceedings of the 69th ICOLD annual meeting. <http://www.stump.de/files/veroeffentlichungen/95.pdf>
- Fiber. [www.ddp13fiberoptics.wordpress.com](http://www.ddp13fiberoptics.wordpress.com). Accessed 28 May 2018
- Grosswig S, Hurtig E, Kühn K, Rudolph F (2001) Distributed fibre-optic temperature sensing technique (DTS) for surveying underground gas storage facilities. Oil Gas Eur Mag 4:31–34
- Guideline for Use of Fibre Optic Sensors, COST Action 299 “FIDES” Document, 2009. <https://infoscience.epfl.ch/record/143489/files/COST299%20Guideline%20FOS.pdf>
- Habel WR (2004) Fiber optic Fabry-Perot sensors, applications and reliability aspects. [https://www.fig.net/nottingham/proc/ts\\_06\\_3\\_habel\\_ppt.pdf](https://www.fig.net/nottingham/proc/ts_06_3_habel_ppt.pdf)
- Habel WR (2009) Reliable use of fiber-optic sensors. In: Boller C, Chang F-K, Fujino Y (eds) Encyclopedia of structural health monitoring, vol 5, 1st edn. Wiley, Chichester, pp 2551–2563
- Habel WR, Gutmann T (2005) Embedded quasi-distributed fibre optic sensors for long-term monitoring of 4,500 kN rock anchors in the Eder Gravity dam in Germany. In: Proceedings of the SHMII-2 conference, vol 1. BALKEMA, Taylor & Francis, pp 289–297

- Habel WR, Heidmann G (2013) Electric power stations and transmission networks. In: Czichos H (ed) *Handbook of technical diagnostics fundamentals and application to structures and systems*, 1st edn. Springer, Berlin/Heidelberg, pp 471–504. ISBN 978-3-642-25849-7
- Habel WR, Jeyapalan JK (2018) Benefits of standards for fiber-optic sensors in soil-structure interaction. To be published in *Geotech Eng J SEAGS & AGSSEA* 50(1). ISSN 0046-5828
- Habel WR, Krebber K (2011) Fiber-optic sensor applications in civil and geotechnical engineering. *Photon Sens* 1:268–280. <https://doi.org/10.1007/s13320-011-0011-x>
- Habel WR, Feddersen I, Fitschen C (1999) Embedded quasi-distributed fiber-optic sensors for the long-term monitoring of the grouting area of rock anchors in a large gravity dam. *J Intell Mater Syst Struct* 10:330–339. <https://doi.org/10.1177/1045389X9901000409>
- Habel WR, Röben R, Hüttl R, Kuchejda M (2011) Monitoring of corrosion protection in reinforced concrete structures using an integrated pH optode. In: *Proceedings of SHMII-5 of ISHMII*. <https://www.researchgate.net/directory/publications>
- Hicke K, Krebber K (2017) Towards efficient real-time submarine power cable monitoring using distributed fibre optic acoustic sensors. SPIE 10323:10323–10494. <https://doi.org/10.1117/12.2267474>
- Hussels M, Chruscicki S, Habib A, Krebber K (2016) Distributed acoustic fibre optic sensors for condition monitoring of pipelines. In: *Proceedings of EWOFS'2016*. <https://doi.org/10.1117/12.2236809>
- IEC 61757:2018. Fiber optic sensors – generic specification
- IEC 61757-1-1 (2016) Strain measurement – strain sensors based on fiber Bragg gratings
- IEC 61757-2-2 (2016) Temperature measurement – distributed sensing
- CD IEC 61757-2-1 (under development) Temperature sensors based on fiber Bragg gratings
- I-MON Interrogation Monitors for FBG sensing systems. <https://ibsen.com/products/interrogation-monitors/>. Accessed 29 May 2018
- intelligent Distributed Acoustic Sensor (iDAST<sup>TM</sup>). <https://silixa.com/technology/idas/>. Accessed 28 May 2018
- Johannessen K, Drakeley BK, Farhadiroshan M (2012) Distributed acoustic sensing – a new way of listening to your well/reservoir. In: *Proceedings of SPE intelligent energy international*. Society of Petroleum Engineers. <https://doi.org/10.2118/149602-MS>
- Ju J, Jin W (2009) Photonic crystal fiber sensors for strain and temperature measurement. *J Sens* 2009:Article ID 476267, 10 pp. <https://doi.org/10.1155/2009/476267>
- Krohn D, Mendez A (2017) Fiber optics sensors standards report. IEEE SA Industry Connections White Paper. IEEE
- Liehr S, Lenke P, Krebber K, Seeger M, Thiele E, Metschies H, Gebreselassie B, Münich JC, Stempniewski L (2008) Distributed strain measurement with polymer optical fibers integrated into multifunctional geotextiles. SPIE 700302. <https://doi.org/10.1117/12.780508>
- Liehr S, Breithaupt M, Krebber K (2017) Distributed humidity sensing in PMMA optical fibers at 500 nm and 650 nm wavelengths. *Sensors* 17:738, 12 pp. <https://doi.org/10.3390/s17040738>
- Malik Y-H, Kölling M, Gräf T, Menge M (2017) Monitoring of partial discharges through fiberoptic sensors in medium voltage switchgear. In: *Proceedings of 20th international symposium on high voltage engineering ISH*, Buenos Aires
- Michlmayr G, Chalari A, Clarke A, Or D (2017) Fiber-optic high-resolution acoustic emission (AE) monitoring of slope failure. *Landslides* 14(3):1139–1146. <https://doi.org/10.1007/s10346-016-0776-5>
- Micron Optics Inc. Applications overview. <http://www.micronoptics.com/applications/>. Accessed 28 May 2018
- Micron Optics sm125 Interrogator. <http://www.micronoptics.com>. Accessed 29 May 2018
- NKT Photonics – LIOS Sensing. <https://www.nktphotronics.com/lios/>. Accessed 28 May 2018
- OZ Optics Ltd. Fiber optic distributed strain and temperature sensors. [http://www.ozoptics.com/products/fiber\\_optic\\_distributed.html](http://www.ozoptics.com/products/fiber_optic_distributed.html). Accessed 28 May 2018
- Propst A, Peters KJ, Zikry MA, Schultz S, Kunzler W, Zhu Z, Wirthlin M, Selfridge R (2010) Assessment of damage in composite laminates through dynamic, full-spectral interrogation of

- fiber Bragg grating sensors. *Smart Mater Struct* 19:015016, 11 pp. <https://doi.org/10.1088/0964-1726/19/1/015016>
- Rohwetter P, Habel WR (2013) Fibre-optic sensors for partial discharge-generated ultrasound in elastomeric high-voltage insulation materials. *SPIE* 8794:879407-1, 4 pp
- Rohwetter P, Habel WR, Heidmann G, Pepper D (2015) Acoustic emission from DC pre-treeing discharge processes in silicone elastomer. *IEEE Trans Dielectr Electr Insul* 22:52–64
- Rota-Rodrigo S, Pinto AMR, Bravo M, Lopez-Amo M (2013) An in-reflection strain sensing head based on a Hi-Bi photonic crystal fiber. *Sensors* 13:8095–8102. <https://doi.org/10.3390/s130708095>
- Santos JL, Farahi F (eds) (2015) Handbook of optical sensors. CRC Press, Taylor & Francis Group, Boca Raton/London/New York
- Schilder C, Kohlhoff H, Hofmann D, Habel WR (2012) Structure-integrated fibre-optic strain wave sensor for pile testing and monitoring of reinforced concrete piles. In: Proceedings of EWSHM-2012. <http://www.ndt.net/article/ewshm2012/papers/we3c4.pdf>
- Schlüter (Schukar) VG (2010) Entwicklung eines experimentell gestützten Bewertungsverfahrens zur Optimierung und Charakterisierung der Dehnungsübertragung oberflächenapplizierter Faser-Bragg-Gitter-Sensoren (Development of an experimental method for strain transfer optimization and characterization of surface-applied fibre Bragg grating sensors). Dissertation, TU Berlin, BAM-Dissertationsreihe, vol 56. ISBN 978-3-98113346-7-8. urn:nbn:de:kobv:b43-1024
- Schukar VG, Baitinger E, Kusche N, Steinke F, Habel WR (2014) Use of spectral conditions to separate strain and temperature effects in fibre Bragg grating sensors embedded in load-carrying anisotropic laminates. *Exp Mech* 54:421–429. <https://doi.org/10.1007/s11340-013-9773-y>
- Schuler S (2010) Matrixintegrierte faseroptische Sensoren für die experimentelle Bestimmung von Mikroverformungen in zementgebundenen Baustoffen (Matrix-integrated fibre-optic sensors for experimental investigation of microdeformation in cementitious materials). Dissertation, TU Berlin, BAM-Dissertationsreihe, vol 65. ISBN 978-3-9813550-8-6. <https://d-nb.info/112103571X/34>
- Schuler S, Habel WR, Hillemeier B (2008) Embedded fibre optic micro strain sensors for assessment of shrinkage at very early ages. In: Proceedings of 1st international conference on microstructure related durability of cementitious composites, Nanjing, 13–15 Oct 2008
- Schuler S, Hillemeier B, Fuhrland M, Meinel D, Habel WR (2009) Untersuchung betontechnologischer Fragestellungen mit Hilfe eingebetteter flexibler faseroptischer Fabry-Perot Interferometer (Investigations of durability parameters of concrete by means of embedded flexible fiber-optic Fabry-Perot interferometers). *Tech Mess* 11:517–526
- Sensuron LLC/Ensysco. Nondestructive damage evaluation using fiber optic sensing (information sheet, 2016). <http://www.sensuron.com/>. Accessed 28 May 2018
- Silixa. Distributed sensing. <https://silixa.com/resources/what-is-distributed-sensing/>. Accessed 30 May 2018
- Smart Fibres. <https://www.smartfibres.com/technology>. Accessed 28 May 2018
- Tan CH, Shee YG, Yap BK, Adikan FRM (2016) Fiber Bragg grating based sensing system: early corrosion detection for structural health monitoring. *Sensors Actuators A Phys* 246:123–128
- Ukil A, Braendle H, Krippner P (2012) Distributed temperature sensing: review of technology and applications. *IEEE Sensors J* 12:885–892. <https://arxiv.org/ftp/arxiv/papers/1503/1503.06261.pdf>
- Van Hoe B, Oman KG, Van Steenberge G, Stan N, Schultz SM, Peters KJ (2017) High-speed interrogation of multiplexed fiber Bragg gratings with spectral distortion. *IEEE Sensors J* 17:6941–6947
- VDI/VDE 2660 Part 2 (2018) Optical temperature sensor based on fibre Bragg grating – fundamentals, characteristics and sensor testing
- Venugopalan T, Yeo TL, Basedau F, Henke AS, Sun T, Grattan KTV, Habel WR (2009) Evaluation and calibration of FBG-based relative humidity sensor designed for structural health monitoring. *SPIE* 750310. <https://doi.org/10.1117/12.835611>

---

## **Part III**

### **Electromagnetic Techniques**



# Eddy Current Testing

19

Zhenmao Chen, Cherdpong Jomdecha, and Shejuan Xie

## Contents

Introduction .....	646
Short Historical Perspective and Progression of ECT .....	647
Basic Principles of Eddy Current Testing .....	649
Advantages and Limitations of ECT .....	650
Framework and Organization of the Chapter .....	651
Theory of Electromagnetics for ECT Problem .....	651
Basic Equations of Electromagnetic Field .....	651
Skin Effect: Standard Depth of Eddy Current Penetration .....	656
Sensitivity of Eddy Current Testing .....	658
Influence Factors on Detectability in Eddy Current Testing .....	663
Conductivity .....	663
Permeability .....	664
Frequency .....	664
Lift-Off and Fill Factor .....	664
Geometry .....	665
Probe Handling .....	665
Discontinuities .....	666
Numerical Methods for Eddy Current Testing .....	666
The $A - \phi$ Method for Three-Dimensional Eddy Current Analysis .....	667
The Ar Method for Three-Dimensional ECT Analysis .....	673
Design and Optimization of ECT Probes .....	676
Types of Sensing Elements for ECT Probes .....	676
Design of ECT Probes Based on Numerical Analysis .....	683
Simplified Method for Detectability Evaluation of ECT Probes .....	691
Optimal Design of ECT Probes for Crack Detection .....	698
Applications of Advanced Numerical Analysis for ECT .....	703
Forward Analysis Schemes for Fast ECT Signal Simulation .....	704
Techniques for Crack Reconstruction from ECT Signals .....	707
References .....	724

Z. Chen (✉) · C. Jomdecha · S. Xie

State Key Laboratory for Strength and Vibration of Mechanical Structures Shaanxi ERC for NDT and Structural Integrity Evaluation School of Aerospace, Xi'an Jiaotong University, Xian, China  
e-mail: [chenzm@mail.xjtu.edu.cn](mailto:chenzm@mail.xjtu.edu.cn); [cherdpong.jom@stu.xjtu.edu.cn](mailto:cherdpong.jom@stu.xjtu.edu.cn); [xiesj2014@mail.xjtu.edu.cn](mailto:xiesj2014@mail.xjtu.edu.cn)

**Abstract**

In various NDT methods, eddy current testing (ECT) technique is widely used for surface and near surface defect inspection, and characterization of electrical conductive materials. This chapter gives brief introduction of theories and applications of advanced ECT, with emphases on the probe design and numerical simulation methods. The chapter moves from short historical and status reviews of the ECT technique, a basic understanding of ECT principles, to state of the art of the testing method in the first section. As bases of ECT numerical simulation methods, theories of electromagnetics related to the advanced ECT are presented in section “[Theory of Electromagnetics for ECT Problem](#).” The topics include basic equations of the low frequency electromagnetic field, skin effect and standard depth of penetration in ECT, and sensitivity and influence factors in ECT inspection. In section “[Numerical Methods for Eddy Current Testing](#),” numerical methods for the three-dimensional ECT problem are described in terms of the  $A\text{-}\phi$ ,  $A_r$  formulations, and FEM and BEM methods. In addition, the equations for calculating ECT signals from the eddy current field are described based on the Biot-Savart’s law and the reciprocity principle at the end of the section. Due to advancement in ECT probe design and optimization, in section “[Design and Optimization of ECT Probes](#),” typical types of ECT probes and magnetic field sensors are introduced. Later, numerical designs of various ECT probes are presented. Furthermore, a phenomenological strategy based on a simplified relationship between the source magnetic field and the induced eddy current is described for evaluation of crack-probe interaction and detectability of ECT probes. At the end of the section, procedures for optimal design of advanced ECT probes for crack detection are given. In section “[Applications of Advanced Numerical Analysis for ECT](#),” progress in forward and inverse numerical techniques and schemes for simulation of ECT problems are explained in detail. Specific numerical approaches are utilized for the ECT signal simulation and crack profile reconstruction by using a deterministic optimization method, an artificial intelligent method, and stochastic optimization methods. The chapter gives good reference for students and researchers in the field of ECT and computational electromagnetics.

---

**Introduction**

In various nondestructive testing (NDT) methods, eddy current testing (ECT) technique is particularly used for surface and near-surface inspection and characterization of electrical conductive materials. The ECT technique is realized by inducing electrical currents in the material to be inspected and observing the interaction between those currents and the material (Udpa and Moore 2004; Cecco et al. 1986). Eddy currents are induced by the alternating electromagnetic field generated by the excitation coil(s) of the ECT probe and monitored simultaneously by measuring changes of probe responses. Since it is an electromagnetic induction process, direct electrical contact between the probe and the inspection target is not required. In addition to the feature of fast inspection, the testing result of ECT is generally in

the form of electrical signals which enables easier signal processing. Sizing and evaluation of the detected flaws are typically established by comparing between the measured flaw signals and the references. Applications of ECT are numerous and widespread covering manufacturing and in-service inspection such as in nuclear power plants, oil and gas pipelines, aerospace, automobile, and other industries (Udpa and Moore 2004; IAEA 2011; NRC 1997; Hellier 2003). Performing ECT is not only for regular inspection as in other NDT methods but can also be used as a structural health monitoring (SHM) technique. In the progression of material sciences and manufacturing processes, components and structures are becoming more complex, large-scale, and high-performance. In addition, they are designed to be used in extreme environments, such as high-temperature, high-pressure, radioactive, and corrosive environments. NDT of such engineering components and structures is particularly important to ensure their integrity. Currently, quantitative nondestructive evaluation (QNDE) has become an essential means for modern industry at almost all stages in the production or life cycle of key structural components, i.e., in material processing, fabrication, and in-service inspection (Davis 1996; Thompson and Chimenti 1992; Achenbach 2000). QNDE is particularly important for the in-service inspection of high-cost and critical load structures and components whose failure could lead to tragic consequences, even disasters. Therefore, the requirements of advanced NDT and evaluation of material integrity are very important for manufacturing processes and maintenance. In an advanced ECT and evaluation, numerical methods are key tools in the development of signal analysis method and in design, upgrade, and optimization of the ECT probes and instrumentation. As well, numerical strategies for inversion of ECT signals can be used for quantitative flaw reconstruction based on the detection signals. In this chapter, the advanced ECT technique, in terms of principles, probe development, and numerical methods for defect sizing from measured signals, is introduced with typical applications.

## Short Historical Perspective and Progression of ECT

As the main method of electromagnetic NDT, the development of ECT can be traced back to the nineteenth century. In 1831, M. Faraday discovered the electromagnetic induction. He induced current flow in secondary coil by switching a battery on and off. After that, in 1879, D.E. Hughes performed the first recorded experiment of ECT for metal sorting. It was the first prelude to the ECT technology (Hughes 1879). Initially, the extreme sensitivity to material resistivity and magnetic permeability and conditions made ECT difficult and unreliable. It took until 1926 before the first ECT instrument was developed to measure sample thickness (Macmaster 1963). By the end of World War II, further research and improved electronics made industrial inspection possible, and many practical instruments were developed. A major breakthrough came in the 1950s when Dr. F. Forster, the founder of Dr. Foerster Institute, Germany, developed ECT instruments with impedance plane signal displays as a flaw detector (Forster 1959). He provided a theoretical basis for the analysis of the mechanism of ECT and research equipment. These works made it possible to discriminate between different

parameters, though the procedure was still empirical. During the 1960s progress in theoretical and practical applications of the advanced technology took ECT from an empirical to an accepted engineering discipline. During that time, other NDT methods such as ultrasonic testing and radiography became well established and ECT played a secondary role, mainly in the aircraft industry. Recent requirements particularly for heat exchanger tube inspection in the nuclear industry contributed significantly to further development of ECT as a fast, accurate and reproducible NDT technique. After 1980, the practical application of the ECT was first reported in the maintenance and inspection of steam generator tubes in PWR nuclear power plants by many researchers and engineers (Richard 1996).

In recent advancements of the ECT technology and equipment, growth in electronic devices, computation system, and data analyses as a “digital age” has brought transformational change for inspection technologies. Research developments in numerical simulation, which mainly relates to low-frequency electromagnetic field problems, have benefited the improvement of accuracy and validation of the ECT (Dodd 1977; Tai 1971; Auld and Moulder 1999). Numerical simulation of the ECT not only involves the basic theory and methods but also the calculation of nonlinear electromagnetic fields, modeling of anisotropic materials, and modeling of complex flaws and structures with high precision and high efficiency (Sabbagh and Sabbagh 1986; Huang et al. 2012; Chen et al. 2015). With technological advancements in numerical techniques, ECT devices and analysis software provide efficient interpretation of quantitative results. These valuable studies are key for the development of QNDE which plays an important role in the prevention of structural failures. Less experienced inspectors are capable of making critical decisions that previously required the approval of Level II and III certified inspectors. Currently, more specialized techniques are employed for certain applications.

According to different detection principles and excitation techniques, the ECT can be divided into single-frequency eddy current, multifrequency eddy current, pulsed eddy current, remote field eddy current, near field eddy current testing, etc. In classifying NDT methods for the purpose of qualifying and certifying testing personnel, the American Society for Nondestructive Testing (ASNT) classifies all of these techniques under the umbrella of electromagnetic testing (ET) method (SNT-TC-1A 2016).

In order to overcome some of the shortcomings of single-frequency ECT, in 1970 Libby and Wandling proposed the multifrequency eddy current (MFEC) method, which uses excitation current signals at several frequencies at the same time (Libby and Wandling 1970). Due to different penetration depths of the excitation signals at different frequencies in the conductor, MFEC can obtain defect information from multiple depths of the test-piece on the one hand and improve the accuracy of ECT detection on the other hand. Analysis contributes to the elimination of interference factors, such as supporting plates, tube sheets, sediments, dents in the tubes of heat exchangers, and interference noise caused by the cold processing of the tubes, probe lift-off, etc., greatly enhancing the detection capability of the ECT technology. However, MFEC can only provide limited detection information and is sometimes difficult to achieve accurate quantification of defects.

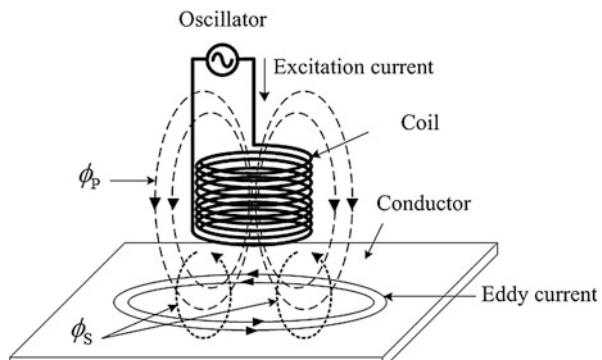
Pulsed eddy current testing (PECT) was developed in recent years to overcome the limited data weakness in MFEC signals (Ludwig and Dai 1990; Xie et al. 2011; Tian et al. 2009). The excitation signal in PECT is in the form of a pulse, usually a square wave with a certain duty cycle, and the response signal contains rich frequency components. Therefore, deeper flaw information of the conductor specimen can be detected. The PECT mainly have the following features: no need to change the parameter settings, one scan can get rich frequency information; a large amount of energy can be applied to the excitation probe to realize detection of deep defects; the cost of instruments is low compared to MFEC. In recent years, PECT technology has achieved considerable progress in the detection of cracks and corrosion in aircraft fuselage structures, etc.

Remote field eddy current testing technology (RFEC) is a technique for eddy current tube inspection by using the far field (or remote field) effect. The probe is generally utilized by transmitting and receiving technique (Rebican et al. 2005; Atherton 1995). A low-frequency alternating current is generated through the excitation coil. The pickup coil can receive the induced magnetic field diffusing back through the tube wall from the eddy currents. Flaw information can be obtained by the changes of pickup coil variables even for flaws outside the tube. In recent years, practical applications and development of RFEC technology have become a hot research topic for inspection and sizing of defects outside the magnetic tube and defects on opposite sides of a magnetic plate.

## Basic Principles of Eddy Current Testing

The principle of ECT is fundamentally the same as that first introduced in 1879. Electromagnetic induction by Oersted's law was used to explain inducing eddy current phenomena. In Fig. 1, the alternating current flowing through the testing coil produces an alternating magnetic field of flux density  $B$ . With more winding in the testing coil, each carrying the same current, the flux density rapidly increases, i.e., the magnetic flux density varies linearly with the current in the testing coil. The

**Fig. 1** Concept diagram of the principle of eddy current testing



total magnetic flux ( $\phi_p$ ) contained within the coil is the product of  $B$  and the area of the coil. When the testing coil is brought near to, or placed on the conductor (specimen), the magnetic field passes into the material and currents are induced in the conductive material. These induced currents are known as eddy current because of their closed circular paths. According to Lenz's law, these induced eddy currents will generate a secondary magnetic field, which opposes the primary magnetic field. The induced magnetic flux ( $\phi_s$ ) also opposes the primary magnetic flux. As the statement of Faraday's law, an electrical current will flow in the conductor when an induced magnetic field cuts a conductor. The alternating current flowing through the testing coil will produce a change of the total magnetic field in the testing coil. When the electrical properties of the material including conductivity and permeability change, they will affect the total induced current in the testing coil. The mutual inductance causes a change in the impedance of the testing coil. The impedance signals sensed by the search coil are the ECT measurements of the specimen. Therefore, using the effect of electromagnetic fields and induction to characterize physical properties and defects in metallic materials is the basic principle of ECT technique.

## Advantages and Limitations of ECT

### Advantages

ECT has significant advantages in NDT by offering a noncontact and high-speed method of inspection. The method is suitable for many industries such as nuclear power plants, oil and gas pipelines, aerospace, automobiles, and other consumer products. The ECT techniques can be utilized for flaw detection and measurements of material properties and specimen dimensions. Portable ECT instruments are available for simple and rapid inspection by individual operators. It can also be embedded with mechanized testing systems to sort many materials, to monitor deterioration of materials and equipment in service, and to verify process quality. During testing or measurement, ECT provides instantaneous responses. Therefore, signal analysis, interpretation, and evaluation of the ECT can be performed in real time. As well, automation systems are compatible with the ECT instrument to develop mass testing of similar parts at high rates and economy. The test results can be optimized for automation of test systems, for sorting of testing parts, for control of manufacturing processes, and for automatic documentation for process control and statistical quality control.

### Limitations

The ECT method is generally limited to application for NDT of conductive materials with significant electrical conductivity such as metals and alloys and composites with conducting layers or reinforcing fibers. The limitation relates to the specific nature of the test, which needs to generate time-varying magnetic fields for exciting eddy current flow into conductive testing materials. The ECT method provides maximum test sensitivity for the surface and near-surface inspection of the test

conductive material adjacent to the source of excitation. In some cases, e.g., in high conductivity materials, eddy current may be difficult to apply to thick specimens because of the skin effect and attenuation of the electromagnetic field at certain depths below the surface. The ECT method also tends to respond to discontinuities that lie transverse to the flow of eddy currents within test materials, where these discontinuities interrupt, or distort the current flow paths. Therefore, it tends to be insensitive to laminar discontinuities, which lie parallel to the induced eddy currents.

## **Framework and Organization of the Chapter**

The chapter moves from the historical perspective and progression reviews of the ECT, a basic understanding of ECT principles, to advantages and limitations of the testing method in this section. Since this chapter focuses on numerical methods, theories of electromagnetics related to the advanced ECT are presented in section “[Theory of Electromagnetics for ECT Problem](#).” The topics include basic equations of the electromagnetic field, skin effect and standard depth of penetration in ECT, and sensitivity and influence factors in ECT inspection. In section “[Numerical Methods for Eddy Current Testing](#),” numerical methods for the three-dimensional eddy current analysis are described in terms of the  $A\text{-}\phi$  method and Ar method. In addition, the numerical strategies for calculating ECT signals from eddy current are described based on Biot-Savart’s law and the reciprocity principle at the end of the section. Due to advancement in ECT probe design and optimization, in section “[Design and Optimization of ECT Probes](#),” typical types of ECT probes and magnetic field sensors are introduced. Later, numerical designs of various ECT probes are presented. Furthermore, a simplified method of phenomenological relationship between the source magnetic field and the induced eddy current is described for evaluation of crack-probe interaction and detectability of ECT probes. At the end of the section, procedures for optimal design of ECT probes for crack detection are discussed. In order to contribute to numerical simulation of ECT, in section “[Applications of Advanced Numerical Analysis for ECT](#),” progress in forward and inverse numerical techniques and strategies for ECT simulation are explained in details. Specific numerical approaches are utilized for the ECT signal simulation and crack shape reconstruction by using a deterministic optimization method, an artificial intelligent method, and stochastic optimization methods.

---

## **Theory of Electromagnetics for ECT Problem**

### **Basic Equations of Electromagnetic Field**

In traditional ECT, the probe is excited by alternating current at a fixed frequency. The maximum excitation frequencies are generally lower than 10 MHz (Dodd and Deeds 1968). Therefore, ECT belongs to the low-frequency electromagnetic field problem. As in physical testing processes, electric and magnetic properties of the test

piece, i.e., flaws, are measured by observing their influences on the detection signal through the probe impedance in the complex impedance plane. This process is ruled by the system of Maxwell's equations for the electromagnetic field. The system equations in differential form and mathematical descriptions for the electromagnetics are Faraday's law of induction, Maxwell-Ampere's law, and Gauss's law in electric and magnetic forms. Supposing all materials within the system are linear and homogeneous, the systems of Maxwell's equations for time-varying fields are (Rothwell and Cloud 2001)

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \text{(Faraday's law of induction)} \quad (1)$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \text{(Maxwell-Ampere's law)} \quad (2)$$

$$\nabla \cdot \mathbf{B} = 0, \text{(Gauss's law for magnetic field)} \quad (3)$$

$$\nabla \cdot \mathbf{J} = 0, \text{(Gauss's law for electric field)} \quad (4)$$

where  $\mathbf{B}$  is the magnetic flux density vector,  $\mathbf{E}$  is the electric field intensity vector,  $\mathbf{H}$  is the magnetic field intensity vector,  $\mathbf{D}$  is the electric flux density vector, and  $\mathbf{J}$  is the eddy current density vector. The electromagnetic phenomena in a closed system can be mathematically solved by using Maxwell's equations with appropriate boundary conditions, analytically or numerically. In addition, Maxwell's equation in terms of constitutive relationships of isotropic material between  $\mathbf{B}$ ,  $\mathbf{H}$ ,  $\mathbf{D}$ , and  $\mathbf{E}$  can be represented as

$$\mathbf{B} = \mu \mathbf{H}, \quad (5)$$

$$\mathbf{J} = \sigma \mathbf{E}, \quad (6)$$

$$\mathbf{D} = \epsilon \mathbf{E}, \quad (7)$$

where  $\mu$  is the magnetic permeability ( $\text{H/m}$ ),  $\epsilon$  is the dielectric constant ( $\text{F/m}$ ), and  $\sigma$  is the electric conductivity ( $\text{S/m}$ ). The permeability of air is usually denoted as  $\mu_0$ .

For the eddy current problem, the dielectric constant of air is  $10^{-9}/36\pi$ , and the electrical conductivity of metallic is usually higher than 1.0 MS/m. Under these circumstances, the term of electric displacement in Eq. (2) can be obtained as  $\partial \mathbf{D} / \partial t = \omega \epsilon / \sigma \mathbf{J} = 0.556 \times 10^{-9} \mathbf{J}$ . The current density due to the electric flux density is much smaller than the induced current and can be ignored based on Eq. (8).

$$\mathbf{J} = \sigma \mathbf{E} \gg \frac{\partial \mathbf{D}}{\partial t}. \quad (8)$$

In fact, the eddy current problem for flaw detection usually uses excitation frequencies below 1 MHz, which corresponds to an electromagnetic wavelength larger than 30 m, which is much larger than the size of the usual probe and detection object.

Therefore, the eddy current problem belongs to the low-frequency electromagnetic field problem. Under this consideration, Eq. (2) can be rewritten as in Eq. (9)

$$\nabla \times \mathbf{H} = \mathbf{J} + \mathbf{J}_0, \quad (9)$$

where the total current density is the sum of eddy current density and the excitation source current density  $\mathbf{J}_0$ .

The eddy current problem also obeys appropriate boundary conditions. The usual boundary conditions are the vertical component of the magnetic flux density and the tangential component of the magnetic field intensity. These conditions are continuous, and when considering them in a magnetic medium, the existence of the magnetic flow at the boundary surface must be considered as

$$\mathbf{B}_1 \cdot \mathbf{n} = \mathbf{B}_2 \cdot \mathbf{n}, \quad (10)$$

$$\mathbf{H}_1 \times \mathbf{n} - \mathbf{H}_2 \times \mathbf{n} = \mathbf{k}, \quad (11)$$

where the subscripts 1 and 2 represent the field on both sides of the boundary medium,  $\mathbf{n}$  is an interface unit normal vector, and  $\mathbf{k}$  is the interfacial surface magnetic current density. In a ferromagnetic medium,  $\mathbf{k} = \mathbf{M} \times \mathbf{n}$ , where  $\mathbf{M}$  is the magnetization vector.

To solve the above Maxwell's system equations, it would be convenient to define a magnetic potential to simplify the system of equations (Dodd and Deeds 1968; Haus and Melcher 1989). Since the magnetostatic fields are nonconservative, a scalar magnetic potential cannot be defined. Fortunately, defining of a vector quantity is satisfactory, which mathematically acts like a magnetic potential. Thus, the magnetic field can be obtained from it by taking a derivative. Denoting the magnetic vector potential as  $\mathbf{A}$ , Gauss's law of magnetics in Eq. (3)  $\nabla \cdot \mathbf{B} = 0$  is used to take advantage of a vector identity that for any vector  $\mathbf{A}$ ,

$$\nabla \cdot (\nabla \times \mathbf{A}) = 0. \quad (12)$$

Therefore, the magnetic flux density  $\mathbf{B}$  can be expressed as the curl of the vector potential  $\mathbf{A}$ ,

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (13)$$

Equation (13) satisfies Gauss's law automatically. Since  $\mathbf{B}$  has units of [Wb/m<sup>2</sup>],  $\mathbf{A}$  has units of [Wb/m]. Meanwhile as  $\mathbf{B} = \mu\mathbf{H}$ , Ampere's law gives

$$\nabla \times \mathbf{B} = \mu\mathbf{J}, \quad (14)$$

$$\nabla \times (\nabla \times \mathbf{A}) = \mu\mathbf{J}. \quad (15)$$

A vector identity of  $\mathbf{A}$  states

$$\nabla^2 \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla \times (\nabla \times \mathbf{A}), \quad (16)$$

where  $\nabla^2 \mathbf{A}$  is the vector Laplacian operator; Eq. (15) can therefore be written as

$$\nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = \mu \mathbf{J}. \quad (17)$$

A gauge condition has to be imposed on vector  $\mathbf{A}$  to obtain a unique solution. In the case of the Coulomb gauge, its divergence needs to be zero, i.e.:

$$\nabla \cdot \mathbf{A} = 0. \quad (18)$$

Finally, Maxwell's governing equation for ECT problem therefore becomes

$$\nabla^2 \mathbf{A} = -\mu \mathbf{J}. \quad (19)$$

A solution to this differential equation is

$$\mathbf{A} = \frac{\mu}{4\pi} \int_V \frac{\mathbf{J}}{R'} dV', \quad (20)$$

where  $R'$  is the distance from the integration point to the point where the field is observed or  $R' = |\mathbf{r} - \mathbf{r}'|$  with  $\mathbf{r}$  is the position vector of the field point and  $\mathbf{r}'$  is the position vector of the source current in the integration. If the current distribution is known,  $\mathbf{A}$  can be computed by using this integral and then obtaining magnetic flux density based on its definition.

Second, one can define a scalar potential by substituting  $\mathbf{B} = \nabla \times \mathbf{A}$  into Eq. (1), which reads

$$\nabla \times \left( \mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = 0. \quad (21)$$

From Eq. (21), the electric field can be written as

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla \phi, \quad (22)$$

with  $\phi$  introduced as the scalar potential to represent the effect of the electric charge. Additionally, substituting Eq. (5) into Eq. (2) and taking into account the definition of the vector potential, the following equation can be derived:

$$\nabla \times \left( \frac{1}{\mu} \nabla \times \mathbf{A} \right) = \mathbf{J}. \quad (23)$$

Substituting Eq. (22) and the constitute relation from Eq. (6) into Eq. (23), we get

$$\nabla \times \left( \frac{1}{\mu} \nabla \times \mathbf{A} \right) = -\sigma \left( \frac{\partial \mathbf{A}}{\partial t} + \nabla \phi \right), \quad (24)$$

In addition, from the continuity condition of eddy currents in Eq. (4), the constitutive relation in Eqs. (6) and (22), we have

$$\nabla \cdot \sigma \left( \frac{\partial \mathbf{A}}{\partial t} + \nabla \phi \right) = 0. \quad (25)$$

Let the magnetic vector potential  $\mathbf{A}$  in Eqs. (24) and (25) satisfy the Coulomb gauge condition as in Eq. (18), i.e.,  $\nabla \cdot \mathbf{A} = 0$ , and introducing the following transformation on the scalar potential  $\phi$ ,

$$\Phi = \int_{-\infty}^t \phi dt, \quad (26)$$

in order that the coefficient matrices are symmetric after the FEM discretization, we can obtain the governing equation using  $\mathbf{A}$  and  $\Phi$  as the following equations in the region of the conducting medium with homogeneous permeability and conductivity. The ECT problem can be described mathematically in the conductor area by the following equation in terms of the magnetic vector potential as (Miya 1995)

$$\frac{1}{\mu} \nabla^2 \mathbf{A} = \sigma \left( \frac{\partial \mathbf{A}}{\partial t} + \nabla \frac{\partial \Phi}{\partial t} \right), \quad (27)$$

$$\nabla \cdot \sigma \left( \frac{\partial \mathbf{A}}{\partial t} + \nabla \frac{\partial \Phi}{\partial t} \right) = 0. \quad (28)$$

Similarly, as the excitation coil of current density  $\mathbf{J}_0$  is in air, the governing equations in the air region become

$$\frac{1}{\mu_0} \nabla^2 \mathbf{A} = -\mathbf{J}_0. \quad (29)$$

The magnetic vector potential  $\mathbf{A}$  can be found by solving Eqs. (27) and (28) using analytical or numerical methods with appropriate boundary conditions imposed by the studied configuration. Once the potentials are obtained, other physical parameters can be deduced from them. The pickup signals can also be obtained from the eddy current calculated above by using Biot-Savart's law. In case of the mutual induction probe (pickup coil and excitation coil are separate), according to Faraday's law, the electromotive force generated in the pickup coil is

$$\Delta V = \sum_{i=1}^N j\omega \int_{S_i} \mathbf{B}_e^f \cdot d\mathbf{S}, \quad (30)$$

where  $B'_e$  is the magnetic flux density generated by the eddy current perturbation at the position of the pickup coil,  $S_i$  is the area enclosed by the  $i$ -th turn of the pickup coil, and  $N$  is the number of turns of the pickup coil. Substituting  $\mathbf{B}'_e = \nabla \times \mathbf{A}'_e$  into Eq. (30) and applying Biot-Savart's law give

$$\Delta V = \sum_{i=1}^N j\omega \oint_{\Gamma_i} \mathbf{A}'_e \cdot d\mathbf{l} = \sum_{i=1}^N j\omega \oint_{\Gamma_i} \left[ \iiint_{\text{cond}} \frac{\mathbf{J}'_e}{R} dv \right] \cdot d\mathbf{l}. \quad (31)$$

For the self-induction probe (pickup coil and excitation coil are the same) with excitation current  $I_0$  (current in every turn), the impedance change of the coil can be obtained by dividing Eq. (31) by the excitation current

$$\Delta Z = \frac{1}{I_0} \sum_{i=1}^N j\omega \oint_{\Gamma_i} \mathbf{A}'_e \cdot d\mathbf{l}. \quad (32)$$

## **Skin Effect: Standard Depth of Eddy Current Penetration**

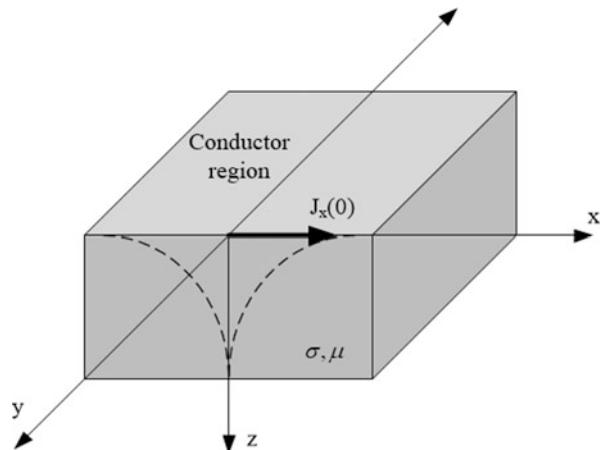
In eddy current problem, eddy currents induced by a changing magnetic field of the excitation coil concentrate near the surface adjacent to the coil. As a time-varying current, it is not distributed uniformly inside the cross section of the specimen but decays exponentially in the depth direction. The decay of eddy current distribution depends on the testing frequency and is a function of the electrical conductivity and magnetic permeability of the specimen. This phenomenon is known as the skin effect (Hayt 2006). Intensity of eddy currents is distributed both parallel to the surface of specimen and decays in depth from the surface of the specimen. In lateral distribution, eddy currents are at a maximum at the surface directly beneath the EC probe and decrease rapidly with increasing lateral distance from the surface. In practice, the eddy current in the specimen at any depth produces magnetic fields which oppose the primary field, thus reducing the net magnetic flux causing a decrease in eddy currents as depth increases. The skin effect is also governed by Maxwell's equations (Eqs. 1, 2, 3, and 4).

In Fig. 2, the current density  $\mathbf{J}$  in a semi-infinite conductor plate of isotropic conductive material is supposed to be induced at the surface by an infinite large current sheet parallel to the plate surface, i.e., to consider a problem in which variables only change in the plate depth direction. In such a case, the eddy current flows only in the  $x$  direction with a density  $\mathbf{J} = J_x \mathbf{u}_x$ . For this one-dimensional electromagnetic field problem, the magnetic field is along  $y$  direction, i.e., only the  $B_y$  component exists. In this case, the governing equations of (1), (2), (3), and (4) become

$$\frac{dJ_x}{dz} = -j\omega\sigma B_y, \quad (33)$$

$$-\frac{dB_y}{dz} = \mu J_x, \quad (34)$$

**Fig. 2** The amplitude variation of the current density vector in the  $z$  direction (thickness direction) from the boundary surface in a homogeneous conductor region



From the equations above, the governing equation for eddy current  $J_x$  can be obtained as

$$\frac{d^2 J_x}{dz^2} = j\omega\mu\sigma J_x, \quad (35)$$

The analytical solution to the above equation has a general form

$$J_x(z,t) = J_{0x} e^{-\beta} \sin(\omega t - \beta), \quad (36)$$

where  $e$  is the base of natural logarithms and  $\beta$  is given by  $z/\delta$  and

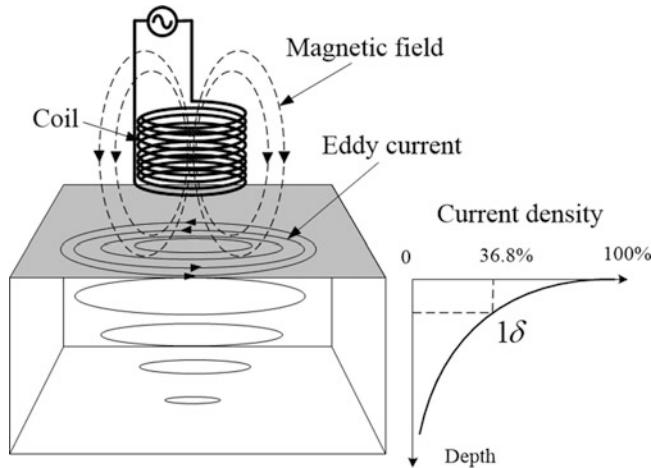
$$\delta = \frac{1}{\sqrt{\pi f \mu \sigma}} \cdot [m] \quad (37)$$

From Eqs. (36) and (37), one finds that the current density decreases more rapidly at higher frequencies and for higher conductivity and permeability. The depth at which the amplitude of  $J_x$  decays exponentially toward the interior of the conducting media, falling to  $1/e$  (or about 36.8%) of the value at the surface, is called the standard depth of eddy current penetration (Ramos et al. 2009) and denoted as  $\delta$  which is described in Fig. 3.

As shown in Eq. (37),  $\delta$  depends on the operating frequency  $f$ , the conductivity  $\sigma$ , and the permeability  $\mu$  of the material. The preceding analysis is approximately valid for a current density penetration in any conductive material whose radius of curvature is much larger than the skin depth. From Eq. (36), the expression can be separated into two components:

$$J_{x1}(z) = J_{0x} e^{-z/\delta}, \quad (38)$$

which describes the exponential decrease in eddy current density in depth, and



**Fig. 3** Standard depth of penetration in eddy current testing

$$J_{x2}(z,t) = \sin(\omega t - z/\delta), \quad (39)$$

which denotes the phase lag of the periodic signal with depth.

Skin effect is responsible as well for lag in phase with increasing depth. The phase lag  $\beta$  also depends on electrical conductivity, permeability, and testing frequency and is given by (Harvey 1995)

$$\beta = \frac{z}{\delta} [\text{radians}] \quad (40)$$

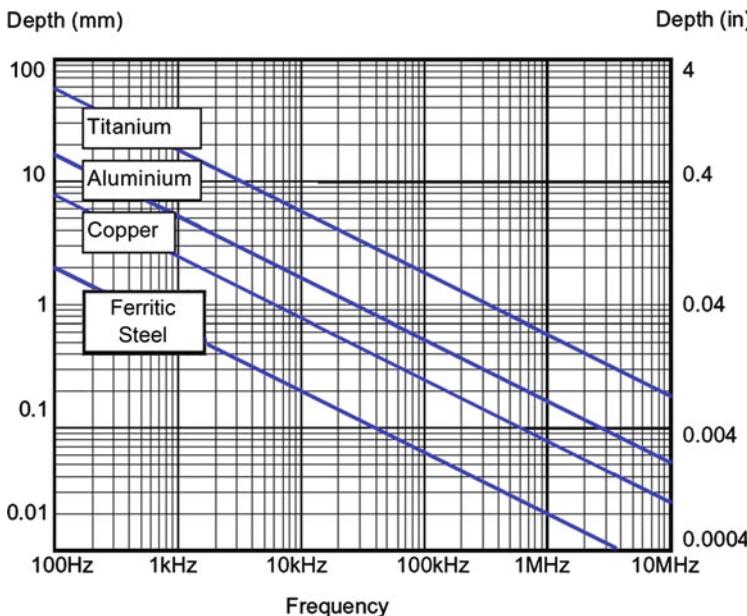
or

$$\beta = \frac{z}{\delta} \cdot \frac{180}{\pi} [\text{degree}] \quad (41)$$

As either the frequency  $f$  or the conductivity  $\sigma$  tends toward infinity, the current density tends to concentrate near the surface of the conductive specimen. Consequently, to examine a deeper thickness or highly conducting or high permeability magnetic materials, lower frequencies are preferred. Typical penetration depths of eddy current for various materials are represented in Fig. 4.

## Sensitivity of Eddy Current Testing

There are two main factors that affect the sensitivity of an ECT system, i.e., the skin depth effect and the test frequency, which causes eddy currents to attenuate with depth depending on the electromagnetic material properties. Therefore, sensitivity of the ECT signals to flaws depends on the current density at the location of the flaw.

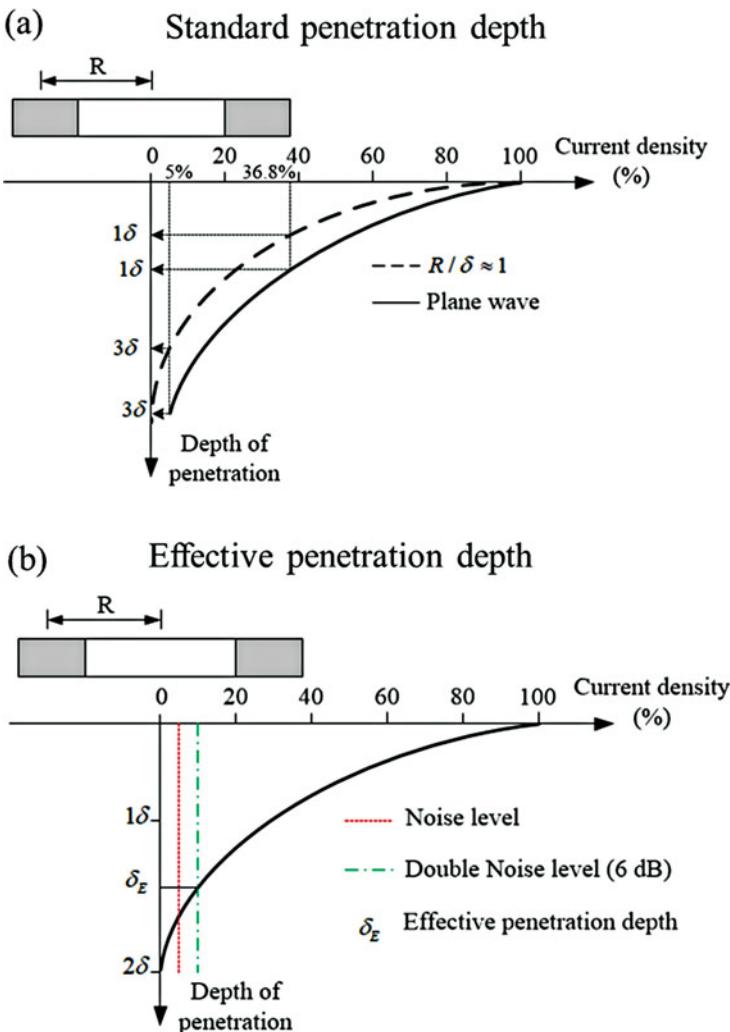


**Fig. 4** Penetration depth of eddy currents for various material and frequencies

Although eddy currents penetrate deeper than one skin depth, they decay exponentially with location in the depth direction. At a position of double the skin depth, i.e.,  $2\delta$ , the amplitude of the eddy current density decays to  $(1/e^2)$  or 13.5% of the surface density. At position of three times of skin depth, the eddy current density drops to only about 5% of the surface density. The sensitivity to subsurface flaw also depends on the eddy current density at that subsurface. It is therefore important to know the effective penetration depth of eddy currents.

In theory, eddy current penetration is analytically computed based on the magnetic field of a plane wave behavior. However, practical ECT probes are far from providing a uniform magnetic field in the air region. For an air core coil-type probe, it is found that the decay of eddy current density strongly depends on the probe diameter (Hagemaier 1985; Mottl 1990). Probes with small diameters ( $\text{radius}/\delta < 1$ ) result in a penetration depth significantly smaller than that of the plane wave as shown in Fig. 5. Although a larger probe diameter can provide a relatively deeper penetration depth of eddy currents, it reduces the sensitivity of the probe for detection of small and short flaws.

Generally, the effective depth of penetration is arbitrarily defined as the depth at which eddy current density decays to 5% of the surface density. For large probes and thick specimen, this depth is about three times the standard skin depth. Unfortunately, for most structural components and practical probe sizes, the effective skin depth is less than  $3\delta$ . Alternatively, the effective depth of penetration can be defined as the depth from where eddy current signals can be received with a sufficient signal-to-noise ratio (Mook and Guericke 2007). Obviously, this effective depth cannot be calculated in general but depends on the material and the defect to be detected, the instrument and



**Fig. 5** Comparison between (a) standard penetration depth and (b) effective penetration depth

probe parameters, and the noise from the environment. Figure 5 illustrates the effective penetration depth for a signal-to-noise ratio of 6 dB as an example. Usually, the effective depth is much greater than the calculated standard skin depth.

### Methods to Enhance the Effective Skin Depth

#### Decrease of the Testing Frequency

Referring to the basic definition of the standard skin depth given in the previous section, the testing frequency is one of the major factors affecting the penetration

depth. It is also the easiest way to increase the effective penetration depth by controlling this factor. Figure 6 shows the results of the decreasing density of the eddy currents for different frequencies. It can be seen that the penetration increases with lowering frequency. Low test frequencies seem to be appropriated for subsurface flaw detection. However, the absolute value of the eddy current density should be viewed as a function of the magnetic field strength. Decreasing test frequency results in lower absolute eddy current density, due to the lower rate of magnetic field strength.

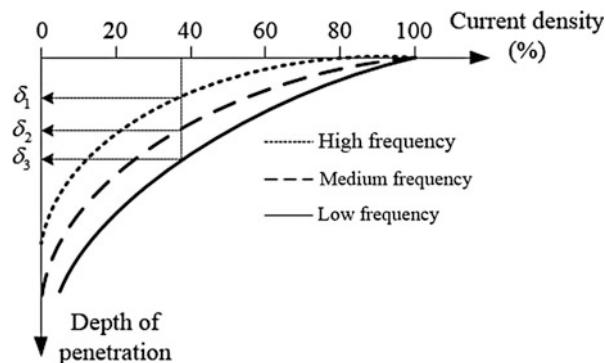
### Increase of the Excitation Field Strength

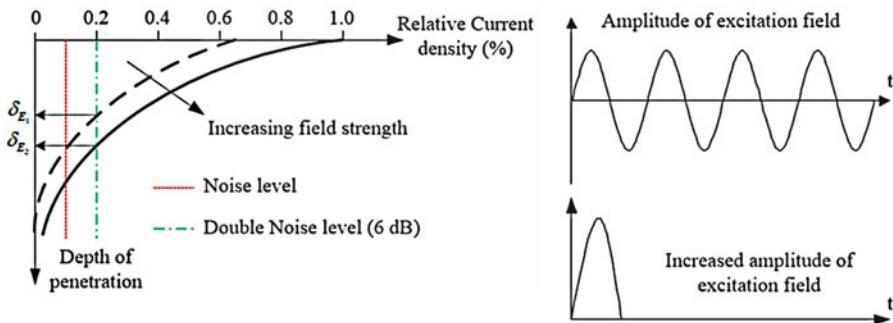
There are two techniques to increase the magnetic field strength, i.e., to increase the excitation current in the excitation probe and to apply pulsed excitation (Janousek et al. 2005; Tian et al. 2009; Chen et al. 2011; Li et al. 2013). The first strategy needs to balance the probe size and heat dissipation. The material properties of the conductor wire used for coil winding such as the electric, magnetic, and thermal properties have to be considered in the design procedure. The latter technique uses a transient excitation current that contains a wide range of excitation frequencies. Because the pulsed eddy current contains low-frequency components, it can also increase the effective depth of penetration. For the PECT method, a typical probe output waveform is as shown in Fig. 7. The surface and subsurface flaws have more influence on the first half of the transient signal, with less influence on the latter half section.

### Using Transmitter and Receiver Technique

Excitation and pickup coils can be designed and optimized by adjusting the distance between the transmitter and receiver coils as shown in Fig. 8 (Fukutomi et al. 1998; Chen et al. 2004; Chen et al. 2005). The excitation coil carries periodic current to produce a magnetic field into the conductive specimen. The eddy currents, in turn, produce their own magnetic field in opposite direction of the magnetic field from the excitation coil. The pickup coil is placed at a distance where the magnetic field from the perturbed eddy currents is dominant. In other words, the pickup coil picks up more information from eddy currents which have penetrated deeper into the material. By monitoring the consistency of the voltage induced in the pickup coil, one can monitor changes or flaws in the specimen.

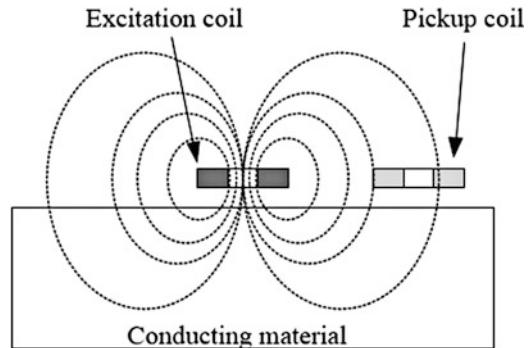
**Fig. 6** Correlation of testing frequency and penetration depth of eddy currents





**Fig. 7** Eddy current testing field strength and penetration

**Fig. 8** Transmit and receive technique

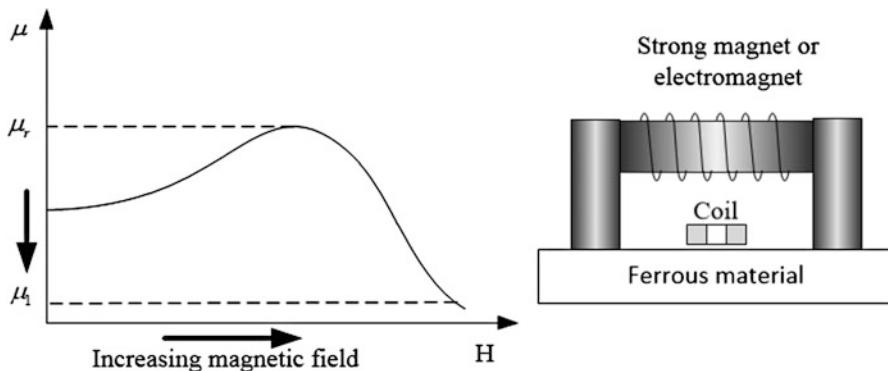


### Changing Material Properties of the Specimen under Test

In the case of a ferromagnetic material, reducing its relative permeability in the way illustrated in Fig. 9 can increase the effective depth of penetration (Higashi et al. 2008; Chen et al. 2014). The magnetic permeability can be reduced by superimposing a strong magnetic field on the material. Application of the strong magnetic field can be realized by using permanent magnets or direct current electromagnets. In theory, the relative permeability can be reduced to 1 if the material is fully saturated. Consequently, the ferromagnetic material under test is transformed to a nonmagnetic material from the eddy current point of view.

### Increasing Sensor Sensitivity at Lower Frequencies

In ECT, detection signals at low frequencies provide the most information on the flaw. Therefore, sensors designed for low-frequency excitation can improve the effective penetration depth in order to detect subsurface flaw in thick materials. These sensor solutions are achieved by using optimized inductive coils or using magnetic field sensors for pickup signals (Hernandez et al. 2012; He et al. 2010). On the other hand, most noise sources in ECT are of high-frequency components such as



**Fig. 9** Reducing material permeability by superimposing a strong magnetic field

the exciting signal, the sensor vibration, and the electronic circuits. Reducing high-frequency noise is also important, which can be done by improving the instrumentation system and adopting digital filtering.

### Influence Factors on Detectability in Eddy Current Testing

During ECT inspection, one must be aware of limitations of the technique and should take maximum advantage of its potential. As is well known, the ECT technique has excellent sensitivity to detect surface breaking flaws. However, sensitivity to deep subsurface flaws is quite limited. Some other factors of ECT systems, such as the probe and scanner, also affect the ECT inspection signals. Therefore, the indication in the inspection signals needs to be correctly interpreted. There are seven factors, as described in the following subsections, influencing this limitation (Udpa and Moore 2004; IAEA 2011).

### Conductivity

Electrical conductivity is a fundamental electromagnetic property that quantifies how a material conducts electric current. Normally, the conductivity of a material can be simplified as homogenous for a whole conductive material, though there are internal material properties that can cause local changes in the conductivity.

### Alloys

Alloys are defined by a metallic bonding character, which are combinations of other metals and/or chemical elements with a base metal. The conductivity of the base metal is affected by each metal and chemical element related to the composition of the alloy. In the ECT, the alloys result from decreased magnetic flux coupling to the metallic test-piece. Therefore, basic metals are often identified by using the ECT to measure the conductivities.

## **Hardness**

Hardness is a measure of the resistance to localized plastic deformation induced by either mechanical manufacturing processes or heat treatments. Hardness changing of metal affects the conductivity of the material. The ECT can be used to classify the hardness levels of the material surface as the qualitative method.

## **Residual Stresses**

Residual stresses are stresses that remain in a solid material after the original cause of the stresses has been removed. Residual stresses which affect to the conductivity changes can occur through a variety of mechanisms including plastic deformations, heat treatment, or structural changes. Local conductivity changes of the material caused by residual stresses can be detected by the ECT by relative measurement with limited depth.

## **Permeability**

The magnetic permeability is the measure of the ability of a material to support the formation of a magnetic field. The permeability constant,  $\mu_0$ , also known as the permeability of free space, is a measure of the amount of reluctance encountered when forming a magnetic field in a vacuum. In ECT, the magnetic field is greatly intensified by the magnetic properties of the material when an excited ECT coil is placed on a ferromagnetic material. If the magnetic field strength at various locations varies even slightly, these small variations have a large effect on the impedance of the ECT coil. These changes in the coil impedance are often so large in comparison to the changes caused by flaws that they could mask all other changes.

## **Frequency**

Normally, material properties of the metal specimen are fixed and the ECT probe is often limited by the material geometry and probe availability. Thus, frequency is an important factor in the ECT that can be controlled by the operator. The main aim of frequency adjustment is to control the depth of penetration, density, and phase of induced eddy currents. In general, higher frequencies are used to detect surface breaking defects, and lower frequencies are used for inspection of subsurface defects. In some cases, multifrequency ECT is required to discriminate between the defect signals and insignificance signals.

## **Lift-Off and Fill Factor**

Lift-off and fill factor are the terms used to describe any space that occurs between the specimen under test and the ECT probe. The lift-off refers to surface coils, and the fill factor refers to encircling and internal coils for tube or bars. Each has an identical effect

on the eddy currents. Decreasing lift-off or fill factor distance affects to enhance the eddy current density on the specimen. On the other hand, increasing lift-off or fill factor distance will decrease eddy current density, which means the testing sensitivity is reduced. Generally, the lift-off or fill factor distance should be kept constant during flaw detection application using the ECT. One benefit of using the effect of lift-off variation signals is to measure the thickness of nonmetallic coatings on metal substrates.

## Geometry

The two main factors in component geometry affecting eddy currents are thickness and edge/end effect.

### Thickness

Changes in thickness of conductive material affect to the ECT if the specimen thickness is less than the effective depth of the eddy current penetration. The ECT is often used to measure the amount of metal thinning caused by manufactured geometry or in-service corrosion/erosion when the inspection access is limited to only one side. Typically, the frequency is selected that produces about one standard depth of penetration at the maximum specimen thickness.

### Edge/End Effect

Eddy currents are distorted when the end or an edge of the inspection specimen is approached with the testing coil since the eddy currents cannot flow outside the specimen. The distortion result produced by the ECT signal when a surface probe approaches the specimen edge is known as “edge effect.” On the other hand, the ECT signal obtained when an internal or encircling probe approaches the end of a tube or rod is known as “end effect.” These signals are very large and will mask any flaw signal at the edge/end area. The limited distance of applying the ECT probe on specimen should be verified to avoid the edge/end effect.

## Probe Handling

As ideal conditions of probe handling, the ECT probe should be scanned at a constant angle to the surface and with a constant lift-off and pressure. Any changes in probe handling may cause significant changes in the ECT signals. In practical inspection using hand held ECT instruments, the effects of probe handling can be reduced with the use of special spring-loaded probes which maintain the probe at a constant angle and pressure to the surface. When scanning close to changes of sections or edge of the specimen, the use of probe guides may assist in good probe handling resulting in a more effective inspection. In addition, the automated scanning systems can be used to improve the quality of the ECT signal, and to construct the inspection image of the scanned area. As with all automated scanning systems, operator variables, such as wobble of the probe, are entirely reduced.

## Discontinuities

Material flaws such as cracks or corrosion or under the ECT perturb the eddy current field due to the fact that the eddy currents must flow around them because crack regions are almost nonconductive. This results in a larger resistance path and a corresponding reduction in eddy current strength. In ideal case, it is essential that the eddy current flow should be preferably perpendicular to the flaw to obtain the maximum ECT sensitivity. If eddy current flow is parallel to the flaw, there will be little perturbation of the currents. In ECT, the use of reference standards in setting up the equipment is particularly important since signals are affected by many different variables. To ensure that inspections are carried out to a repetitive standard, reference blocks with artificial defects are necessary. The reference blocks should be of similar material specification (alloying, heat treatment, conductivity) to the component under test. By setting up the flaw detector (standardization) to give a known response from the artificial defect, an inspection can be carried out repeatedly on the same standard test block. The flaw detector sensitivity settings (standardization) must be checked at regular intervals.

---

## Numerical Methods for Eddy Current Testing

Analyses of eddy current of simple ECT problems are often done empirically for probe development (Fuller 2006; Chen et al. 2008; Yusa et al. 2007a), as prediction of ECT signals is indispensable for detectability evaluation of a new probe configuration. Unfortunately, those empirical methods usually lack efficiency and accuracy and may consume much cost and time to cope with complex testing conditions of various applications. In contrast, analyses using numerical methods are much more efficient as well as time saving than the experimental trial and error approach. Nowadays, more and more resources of numerical analysis methods and codes are developed for NDT applications. Numerical analyses of ECT problems with high accuracy and efficiency have also been realized based on the electromagnetic field theory and numerical methods. As the prediction of ECT signals through numerical modeling is more efficient and reliable compared with experimental methods (Zenglu et al. 2011; Huang et al. 2003; Yusa et al. 2007b), reliable numerical methods are key tools for the development of advanced ECT probes, instruments, and signal processing algorithms.

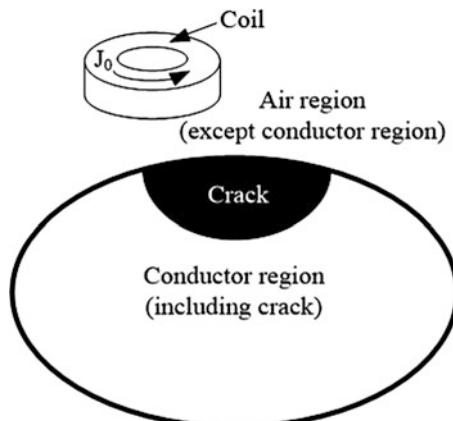
The numerical approach is based on a discretization of the analysis region and on an approximation of solutions. Three methods were adopted for the numerical simulation of ECT problems, i.e., the finite element method (FEM), the boundary element method (BEM), and the finite difference method (FDM) (Chari 1974; Demerdash and Nehl 1978; Ishibashi 1995). In general, analytical solution of ECT problems is difficult except in some very special cases of canonical geometry (plate, tube) and exciting coils with regular shape (Dodd and Deeds 1968; Tegopoulos and Kriegis 1985; Burke 1986). Numerical solutions to three-dimensional ECT problems have been successfully realized and applied to many

practical applications. Several numerical codes using different formulation and different discretization methods have been developed and validated, and some benchmark problems have been proposed based on these codes (Takagi et al. 1994, 1997b; JSATEM 1997) for applications to ECT in nuclear power plants. As two typical numerical codes, the formulations of  $A - \phi$  and  $A_r$  method were solved with an FEM-BEM hybrid strategy and full FEM for numerical analysis of ECT problems (Miya 1995; Chen and Miya 1998b; Chen et al. 1999, 2006, 2009). In the following sections, the formulae and discretization procedures of these two numerical methods are described in detail in addition to the formulation for pickup signal calculations.

### The $A - \phi$ Method for Three-Dimensional Eddy Current Analysis

In FEM discretization for ECT problems, air is a generally coupling medium between the probe and the conductor to be inspected. Thus, the entire calculation area including the air area needs to be subdivided into finite elements. Since the ECT probes may have complicated geometries and a scan is necessary for ECT inspection, the conventional FEM is not effective for simulation of ECT signals. Aiming to solve the  $A - \phi$  formulae given in section “[Theory of Electromagnetics for ECT Problem](#)” of this chapter and numerical model shown in Fig. 10, a hybrid strategy of FEM-BEM is utilized to predict the scanning ECT signals with fast computational speed and reduced memory resources. The FEM-BEM hybrid method uses finite element discretization for the conductor region, and the influence of the excitation current and air region is considered by using the boundary element method, i.e., the air area does not need to be subdivided with finite elements. In this way, the hybrid approach is suitable for eddy current field calculations for complex probes and scanning paths as the probes do not need to be meshed. The detailed equations and discretization processes of the hybrid FEM-BEM method are described in this

**Fig. 10** Numerical model for the  $A - \phi$  method



section based on Galerkin's finite element approach. In practice, based on the governing equations of ECT problem given in section “[Theory of Electromagnetics for ECT Problem](#),” Eqs. (27) and (28) are discretized with FEM, and Eq. (29) is treated with BEM in order to localize the analysis region to the conductor to be inspected. The detailed procedures are as follows:

### Finite Element Discretization in Conductor Region

To discretize Eq. (27), we take the  $x$  component of the vector potential  $\mathbf{A}$  as an example. From Eq. (27), the scalar equation corresponding to the  $A_x$  component is as follows:

$$\frac{1}{\mu_0} \nabla^2 A_x = \sigma \left( \frac{\partial A_x}{\partial t} + \frac{\partial}{\partial t} \frac{\partial \Phi}{\partial x} \right). \quad (42)$$

Multiplying the variation  $\delta A_x$  with Eq. (42) and integrating it over the whole conductor volume  $\Omega$ , the left hand of Eq.(42) becomes

$$\int_{\Omega} \delta A_x \frac{1}{\mu_0} \nabla^2 A_x d\Omega = \frac{1}{\mu_0} \int_{\Gamma} \delta A_x \frac{\partial A_x}{\partial n} d\Gamma - \frac{1}{\mu_0} \int_{\Omega} \nabla(\delta A_x) \cdot \nabla A_x d\Omega. \quad (43)$$

Upon subdividing the conductor into finite elements and interpolating the potentials  $A_x$  with the shape function matrix  $[N]$  as  $A_x(\mathbf{r}) = [N(\mathbf{r})] \{A_x\}$ , the two terms of the right hand of Eq. (43) change to the following expressions:

$$\frac{1}{\mu_0} \int_{\Gamma} \delta A_x \frac{\partial A_x}{\partial n} d\Gamma = \{\delta A_x\}^T \frac{1}{\mu_0} \int_{\Gamma} \{N\}^T \{N\} d\Gamma \left\{ \frac{\partial A_x}{\partial n} \right\}, \quad (44)$$

$$\begin{aligned} & \frac{1}{\mu_0} \int_{\Omega} \nabla(\delta A_x) \cdot \nabla A_x d\Omega \\ &= \{\delta A_x\}^T \frac{1}{\mu_0} \int_{\Omega} \left\{ \frac{\partial[N]^T}{\partial x} \frac{\partial[N]}{\partial x} + \frac{\partial[N]^T}{\partial y} \frac{\partial[N]}{\partial y} + \frac{\partial[N]^T}{\partial z} \frac{\partial[N]}{\partial z} \right\} d\Omega \{A_x\}. \end{aligned} \quad (45)$$

Multiplying the variation  $\delta A_x$  with Eq. (42), the right-hand side becomes the following expressions:

$$\sigma \int_{\Omega} \left( \delta A_x \frac{\partial A_x}{\partial t} \right) d\Omega = \sigma \{\delta A_x\}^T \int_{\Omega} [N]^T [N] d\Omega \left\{ \frac{\partial A_x}{\partial t} \right\}, \quad (46)$$

$$\sigma \int_{\Omega} \left( \delta A_x \frac{\partial}{\partial t} \frac{\partial \Phi}{\partial x} \right) d\Omega = \sigma \{\delta A_x\}^T \int_{\Omega} [N]^T \frac{\partial[N]}{\partial x} d\Omega \left\{ \frac{\partial \Phi}{\partial t} \right\}. \quad (47)$$

Discretizing the equations of the  $y$  and  $z$  components for the vector potential  $\mathbf{A}$  in the same way, we obtain the FEM system equations in Eq. (44) through Eq. (47) in a matrix form as follows as  $\delta A_x$  is arbitrary:

$$\begin{bmatrix} N_1 & 0 & 0 & 0 \\ 0 & N_1 & 0 & 0 \\ 0 & 0 & N_1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} A_x \\ A_y \\ A_z \\ \Phi \end{Bmatrix} + \begin{bmatrix} N_2 & 0 & 0 & N_a \\ 0 & N_2 & 0 & N_b \\ 0 & 0 & N_2 & N_c \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} \partial A_x / \partial t \\ \partial A_y / \partial t \\ \partial A_z / \partial t \\ \partial \Phi / \partial t \end{Bmatrix} = \begin{Bmatrix} f_x \\ f_y \\ f_z \\ 0 \end{Bmatrix}, \quad (48)$$

where the submatrices have the following expressions:

$$[N_1] = \frac{1}{\mu_0} \int_{\Omega} \left\{ \frac{\partial [N]^T}{\partial x} \frac{\partial [N]}{\partial x} + \frac{\partial [N]^T}{\partial y} \frac{\partial [N]}{\partial y} + \frac{\partial [N]^T}{\partial z} \frac{\partial [N]}{\partial z} \right\} d\Omega, \quad (49)$$

$$[N_2] = \sigma \int_{\Omega} [N]^T [N] d\Omega, \quad (50)$$

$$[N_a] = \sigma \int_{\Omega} [N]^T \frac{\partial [N]}{\partial x} d\Omega, \quad (51)$$

$$[N_b] = \sigma \int_{\Omega} [N]^T \frac{\partial [N]}{\partial y} d\Omega, \quad (52)$$

$$[N_c] = \sigma \int_{\Omega} [N]^T \frac{\partial [N]}{\partial z} d\Omega, \quad (53)$$

and the elements of the vector in the right-hand side of Eq.(48) represent the effect of the gradients at the conductor surface

$$\{f_x\} = \frac{1}{\mu_0} \int_{\Gamma} [N]^T [N] d\Gamma \left\{ \frac{\partial A_x}{\partial n} \right\}, \quad (54)$$

$$\{f_y\} = \frac{1}{\mu_0} \int_{\Gamma} [N]^T [N] d\Gamma \left\{ \frac{\partial A_y}{\partial n} \right\}, \quad (55)$$

$$\{f_z\} = \frac{1}{\mu_0} \int_{\Gamma} [N]^T [N] d\Gamma \left\{ \frac{\partial A_z}{\partial n} \right\}. \quad (56)$$

Denoting the coefficient matrix in Eqs. (54), (55), and (56) as a distribution matrix  $[M_1]$ , the unknown vector potentials as  $\{A_d\}$ , and the coefficient matrices as  $[P]$  and  $[Q]$ , respectively, Eq. (48) becomes

$$\{f_x, f_y, f_z\}^T = \{f\} = [M] \left\{ \frac{\partial A_d}{\partial n} \right\} = [M] \left\{ \frac{\partial A_x}{\partial n}, \frac{\partial A_y}{\partial n}, \frac{\partial A_z}{\partial n} \right\}^T, \quad (57)$$

where

$$[M] = \begin{bmatrix} [M_1] & 0 & 0 \\ 0 & [M_1] & 0 \\ 0 & 0 & [M_1] \end{bmatrix} \quad (58)$$

and

$$[M_1] = \frac{1}{\mu_0} \int_{\Gamma} [N]^T [N] d\Gamma. \quad (59)$$

Then, Eq. (48) can be rewritten as

$$[P] \begin{Bmatrix} A_d \\ \Phi \end{Bmatrix} + [Q] \begin{Bmatrix} \partial A_d / \partial t \\ \partial \Phi / \partial t \end{Bmatrix} = \begin{Bmatrix} f \\ 0 \end{Bmatrix}, \quad (60)$$

where the coefficient matrices  $[P]$  and  $[Q]$  represent the left-hand side of Eq. (48).

From Eq. (28), by multiplying the variation  $\delta\Phi$  and performing the same discretization process, we obtain

$$\begin{aligned} & \int_{\Omega} \delta\Phi \nabla \cdot \sigma \left( \frac{\partial \mathbf{A}}{\partial t} + \nabla \frac{\partial \Phi}{\partial t} \right) d\Omega \\ &= \int_{\Gamma} \delta\Phi \sigma \left( \frac{\partial \mathbf{A}}{\partial t} + \nabla \frac{\partial \Phi}{\partial t} \right) \cdot d\Gamma - \int_{\Omega} \nabla(\delta\Phi) \cdot \sigma \left( \frac{\partial \mathbf{A}}{\partial t} + \nabla \frac{\partial \Phi}{\partial t} \right) d\Omega. \end{aligned} \quad (61)$$

Since the normal component of electric field strength at the conductor surface does not exist, the surface integral in the above equation is zero. Thus, the discretization result of Eq. (28) becomes

$$\begin{aligned} & \{\delta\Phi\} \int_{\Omega} \left[ \left[ \frac{\partial N_x}{\partial x} \right]^T [N] \{A_x\} + \left[ \frac{\partial N_y}{\partial y} \right]^T [N] \{A_y\} + \left[ \frac{\partial N_z}{\partial z} \right]^T [N] \{A_z\} \right] d\Omega \\ &+ \{\delta\Phi\} \int_{\Omega} [\nabla N]^T \cdot [\nabla N] d\Omega \{\Phi\} = 0. \end{aligned} \quad (62)$$

We can write this in matrix form as

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ N_a^T & N_b^T & N_c^T & N_1' \end{bmatrix} \begin{Bmatrix} \partial A_x / \partial t \\ \partial A_y / \partial t \\ \partial A_z / \partial t \\ \partial \Phi / \partial t \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}, \quad (63)$$

where

$$\begin{bmatrix} N_1' \end{bmatrix} = \sigma \mu [N_1]. \quad (64)$$

Denoting the coefficient matrix as  $[R]$ , the left-hand side of Eq. (63) becomes

$$[R] \begin{Bmatrix} \frac{\partial A_d}{\partial t} \\ \frac{\partial \Phi}{\partial t} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}. \quad (65)$$

Summarizing Eqs. (48) and (63), we have the FEM system equations in the conductor as

$$\begin{bmatrix} N_1 & 0 & 0 & 0 \\ 0 & N_1 & 0 & 0 \\ 0 & 0 & N_1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} A_x \\ A_y \\ A_z \\ \Phi \end{Bmatrix} + \begin{bmatrix} N_2 & 0 & 0 & N_a \\ 0 & N_2 & 0 & N_b \\ 0 & 0 & N_2 & N_c \\ N_a^T & N_b^T & N_c^T & N_1' \end{bmatrix} \begin{Bmatrix} \frac{\partial A_x}{\partial t} \\ \frac{\partial A_y}{\partial t} \\ \frac{\partial A_z}{\partial t} \\ \frac{\partial \Phi}{\partial t} \end{Bmatrix} = \begin{Bmatrix} f_x \\ f_y \\ f_z \\ 0 \end{Bmatrix}. \quad (66)$$

Representation by using the coefficient matrix  $[P]$ ,  $[Q]$ ,  $[R]$  can be rewritten as

$$[P] \begin{Bmatrix} A_d \\ \Phi \end{Bmatrix} + [Q + R] \begin{Bmatrix} \frac{\partial A_d}{\partial t} \\ \frac{\partial \Phi}{\partial t} \end{Bmatrix} = \begin{Bmatrix} f \\ 0 \end{Bmatrix}. \quad (67)$$

Since the ECT problem is time-harmonic, the following formulae can be finally deduced by using the complex approximation of potential variables:

$$[P + j\omega Q + j\omega R] \begin{Bmatrix} A \\ \Phi \end{Bmatrix} = \begin{Bmatrix} f \\ 0 \end{Bmatrix}. \quad (68)$$

The final form of the formula (68) is the time-harmonic eddy current problems in the conductor area through finite element discretization.

### Discretization by Using Boundary Elements in the Air Region

To avoid finite element meshing for the air region and to treat movement of the ECT coil more easily, the BEM approach is a suitable approach for discretization of the governing equations in the air region. For BEM, the discretized scalar equation for the  $A_x$  component can be deduced by multiplying Eq. (29) with the elementary solution  $u^* = 1/4\pi R$  of the Laplace equation ( $R$  is the distance between source and field point) and integrating over the whole space

$$\int_{\Omega} \mu^* (\nabla^2 A_x dV + \mu_0 J_{0x}) d\Omega = 0. \quad (69)$$

Calculations transform Eq. (69) into

$$\begin{aligned} \int_{\Omega} \frac{1}{4\pi R} \nabla \cdot \nabla A_x d\Omega &= \frac{1}{4\pi} \int_{\Gamma} \frac{1}{R} \frac{\partial A_x}{\partial n} d\Gamma - \frac{1}{4\pi} \int_{\Gamma} A_x \nabla \frac{1}{R} \cdot d\Gamma + \frac{1}{4\pi} \int_{\Omega} A_x \nabla \\ &\quad \cdot \nabla \frac{1}{R} d\Omega. \end{aligned} \quad (70)$$

Considering that the electromagnetic field at the infinite boundary of the air region is zero, the integration area of the first two terms of the right hand of Eq. (70) is limited to the conductor surface. On the other hand, as  $\nabla^2(1/R) = \delta(\mathbf{r} - \mathbf{r}')$ , the third term is nonzero

only when the source point coincides with the field point. Equation (69) must be satisfied for every discrete point (field point) on the conductor surface. The third term is 0.5 once the field point is at the conductor surface, 1/4 for the field point at the edge, and 1/8 for the corner field points. Based on Eq. (70), the weak form of Eq. (69) can be written as

$$\frac{1}{\mu_0} C(p) A_x + \frac{1}{4\pi\mu_0} \int_{\Gamma} \frac{1}{R} \frac{\partial A_x}{\partial n} d\Gamma - \frac{1}{4\pi\mu_0} \int_{\Gamma} A_x \nabla \frac{1}{R} \cdot d\Gamma = \int_{\Omega} \frac{J_{0x}}{4\pi R} d\Omega. \quad (71)$$

The above equation holds true for all field points on the surface of the conductor, with  $C(p)$  equal to 1/2, 1/4, and 1/8, respectively, when the field point is at the surface, edge, and corner.

Discretizing Eq. (71) with the shape function matrix  $A_x(\mathbf{r}') = [N(\mathbf{r}')]\{A_x\}$ , we get

$$\frac{1}{\mu_0} C(p) A_{xi} - \frac{1}{4\pi\mu_0} \int_{\Gamma} [N] \nabla \frac{1}{R} \cdot d\Gamma \{A_x\} + \frac{1}{4\pi\mu_0} \int_{\Gamma} [N] \frac{1}{R} d\Gamma \left\{ \frac{\partial A_x}{\partial n} \right\} = \int_{\Omega} \frac{J_{0x}}{4\pi R} d\Omega, \quad (72)$$

where the right-hand term can be denoted as  $F_{0x}^i$  which corresponds to the current source effect at the  $i$ -th field point. Equation (72) is valid for each vector potential component at all boundary element nodes. In this way, the boundary element equations with the same number of surface nodes of the three vector potential components can be obtained. From Eq. (72), denoting the coefficient matrices as  $[H_x]$  and  $[G_x]$ , the equation corresponding to the  $A_x$  component is formed as

$$[H_x]\{A_x\} + [G_x] \left\{ \frac{\partial A_x}{\partial n} \right\} = \{F_{0x}\}. \quad (73)$$

Summarizing the equations for the  $A_x$ ,  $A_y$ , and  $A_z$  components, the discrete BEM system equations can be obtained as

$$\begin{bmatrix} H_x & 0 & 0 \\ 0 & H_y & 0 \\ 0 & 0 & H_z \end{bmatrix} \begin{Bmatrix} A_x \\ A_y \\ A_z \end{Bmatrix} + \begin{bmatrix} G_x & 0 & 0 \\ 0 & G_y & 0 \\ 0 & 0 & G_z \end{bmatrix} \begin{Bmatrix} \partial A_x / \partial n \\ \partial A_y / \partial n \\ \partial A_z / \partial n \end{Bmatrix} = \begin{Bmatrix} F_{0x} \\ F_{0y} \\ F_{0z} \end{Bmatrix} \quad (74)$$

or simply written as

$$[H]\{A_d\} + [G] \left\{ \frac{\partial A_d}{\partial n} \right\} = \{F_0\}. \quad (75)$$

Multiplying Eq. (75) with the inverse matrix  $[G]^{-1}$ , we get

$$[G]^{-1}[H]\{A_d\} + \left\{ \frac{\partial A_d}{\partial n} \right\} = [G]^{-1}\{F_0\}. \quad (76)$$

Further multiplying with the distribution matrix  $[M]$  results in

$$[M][G]^{-1}[H]\{A_d\} + [M]\left\{\frac{\partial A_d}{\partial n}\right\} = [M][G]^{-1}\{F_0\}. \quad (77)$$

By symmetrizing the coefficient matrix, we obtain the following system equations for the fields in free space:

$$[K]\{A_d\} = -[M]\left\{\frac{\partial A_d}{\partial n}\right\} + [M][G]^{-1}\{F_0\}, \quad (78)$$

where

$$[K] = \frac{1}{2} \left( [K_B] + [K_B]^T \right), \quad (79)$$

$$[K_B] = [M][G]^{-1}[H]. \quad (80)$$

Finally, the  $\mathbf{A} - \phi$  formulation discretized with the hybrid FEM-BEM strategy is as follows:

$$[P + j\omega Q + j\omega R + K] \begin{Bmatrix} A_d \\ \Phi \end{Bmatrix} = \begin{Bmatrix} [M][G]^{-1}\{F_0\} \\ 0 \end{Bmatrix}. \quad (81)$$

## The Ar Method for Three-Dimensional ECT Analysis

### Governing Equations of the Ar Method

To reduce unknowns, the edge finite element is an efficient approach to the solution of the governing equations of the  $A_r$  method (reduced magnetic vector potential method). In such a case, a gauge condition that sets the scalar potential to zero is valid. Therefore, only the vector potential needs to be taken as unknown.

As  $\mathbf{J} = \mathbf{J}_s$  and  $\mu = \mu_0$  in the air region, the governing equation in the air can be written as

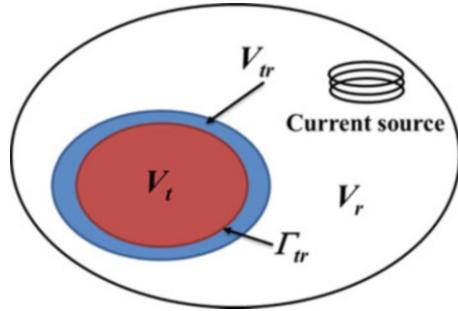
$$\nabla \times \left( \frac{1}{\mu_0} \nabla \times \mathbf{A} \right) = \mathbf{J}_s, (\text{in air}) \quad (82)$$

where  $\mu_0$  is the permeability of the air and  $\mathbf{J}_s$  is the source current density in the excitation coil. On the other hand, the governing equation in conductor can be written as

$$\nabla \times \left( \frac{1}{\mu} \nabla \times \mathbf{A} \right) + \sigma \frac{\partial \mathbf{A}}{\partial t} = 0, (\text{in the conductor}) \quad (83)$$

where  $\mu$  is the permeability of the conductor.

**Fig. 11** Numerical model of the Ar method



In the  $A_r$  formulation, the whole analysis region is divided into two parts as shown in Fig. 11, with  $V_t$  the part including the conductive material where the standard magnetic potential  $\mathbf{A}$  is adopted as unknown and  $V_r$  the analysis region surrounding  $V_t$  where the reduced magnetic vector potential  $\mathbf{A}_r$  is used as unknown ( $\mathbf{A}_r = \mathbf{A} - \mathbf{A}_s$  with  $\mathbf{A}_s$  the magnetic vector potential generated by the source current alone). In addition,  $V_{tr}$  is the edge region between  $V_r$  and  $V_t$ , where  $\mathbf{A}_r$  is also used to denote the unknowns. In practice,  $\mathbf{A}_s$  and  $\mathbf{H}_s$  can be calculated from the excitation coils by using the Biot-Savart's law in free space ( $\mathbf{H}_s$  is the magnetic field intensity generated by the current source and will be used without explanation hereafter)

$$\mathbf{A}_s = \frac{\mu_0}{4\pi} \int_{V_c} \frac{\mathbf{J}_s}{R} dV, \text{ and} \quad (84)$$

$$\mathbf{H}_s = \frac{1}{4\pi} \int_{V_s} \mathbf{J}_s \times \nabla \frac{1}{R} dV, \quad (85)$$

where  $V_c$  is the region of the current source (excitation coil) and  $R$  is the distance from the point of current source  $\mathbf{J}_s$  to point of the field of  $\mathbf{A}_s(\mathbf{H}_s)$ .

The governing equation of Eq. (82) in the air can be derived as

$$\nabla \times \frac{1}{\mu_0} \nabla \times \mathbf{A}_r + \nabla \times \frac{1}{\mu_0} \nabla \times \mathbf{A}_s = \mathbf{J}_s \quad (86)$$

Therefore, the governing equations of the  $A_r$  method can be written as

$$\nabla \times \frac{1}{\mu_0} \nabla \times \mathbf{A}_r = 0, \text{ in } V_r + V_{tr} \quad (87)$$

$$\nabla \times \frac{1}{\mu} \nabla \times \mathbf{A} + \sigma \frac{\partial \mathbf{A}}{\partial t} = 0, \text{ in } V_t \quad (88)$$

where  $\sigma$  is the conductivity of the conductor.

According to the boundary conditions, in joining these two regions, the continuity of the magnetic flux density and magnetic field intensity can be represented by

$$\mathbf{n} \times \mathbf{A} = \mathbf{n} \times (\mathbf{A}_r + \mathbf{A}_s), \text{ on } \Gamma_{tr} \text{ and} \quad (89)$$

$$\mathbf{n} \times \frac{1}{\mu} (\nabla \times \mathbf{A}) = \mathbf{n} \times \left( \frac{1}{\mu_0} \nabla \times \mathbf{A}_r + \mathbf{H}_s \right), \text{ on } \Gamma_{tr} \quad (90)$$

where  $\mathbf{n}$  is a unit normal vector on  $\Gamma_{tr}$  (boundary surrounding  $V_t$ ).

### Discretization Using the Full FEM with Edge Elements

To discretize the governing equations by using the FEM with edge elements, the following Galerkin's finite element formulation is valid. In this case, hexahedral elements with 8 nodes and 12 edges are suitable for linear interpolation of unknowns. In practice, the magnetic vector potential can be discretized as follows for different regions by using the vector shape functions:

$$\mathbf{A} = \sum_{e \in E_t} A^e \mathbf{N}^e, \text{ in } V_t \quad (91)$$

$$\mathbf{A}_r = \sum_{e \in E_r} A^e \mathbf{N}^e, \text{ in } V_r \quad (92)$$

$$\mathbf{A}_r = \sum_{e \in E_r} A^e \mathbf{N}^e - \sum_{e \in E_{tr}} A_s^e \mathbf{N}^e, \text{ in } V_{tr} \quad (93)$$

where  $\mathbf{N}^e$  is the vector shape function of the edge element and  $E_t$ ,  $E_r$ , and  $E_{tr}$  are the sets of edges in  $V_t$ ,  $V_r + V_{tr}$ , and  $\Gamma_{tr}$ , respectively. Consequently, Eqs. (87) and (88) can be discretized as.

$$\begin{aligned} & \int_{V_t} \nabla \times \mathbf{N} \cdot \left( \frac{1}{\mu} \nabla \times \mathbf{A} \right) dV + \int_{V_r + V_{tr}} \nabla \times \mathbf{N} \cdot \left( \frac{1}{\mu_0} \nabla \times \mathbf{A}_r \right) dV \\ & + j\omega \int_{V_c} \mathbf{N} \cdot \sigma \mathbf{A} dV - \oint_{\Gamma_{tr}} (\mathbf{N} \times \mathbf{H}_s) \cdot \mathbf{n} dS = 0 \end{aligned} \quad (94)$$

Equation (94) can be written in matrix form as

$$[K_e] \{\mathbf{A}\} + j\omega [C_e] \{\mathbf{A}\} = [M_e] I, \quad (95)$$

where

$$[K_e] = \frac{1}{\mu} \int_V \nabla \times \{\mathbf{N}\} \cdot \nabla \times \{\mathbf{N}\}^T dV, \quad (96)$$

$$[C_e] = \sigma \int_{V_c} \{\mathbf{N}\} \cdot \{\mathbf{N}\}^T dV, \quad (97)$$

$$[M_e] = \frac{1}{\mu} \{\alpha_s\}^T \int_{V_r} \nabla \times \{N\} \cdot \nabla \times \{N\}^T dV + \oint_{\Gamma_r} (N \times h_s) \cdot n dS \quad (98)$$

with  $I$  denoting the amplitude of the current source.

## Design and Optimization of ECT Probes

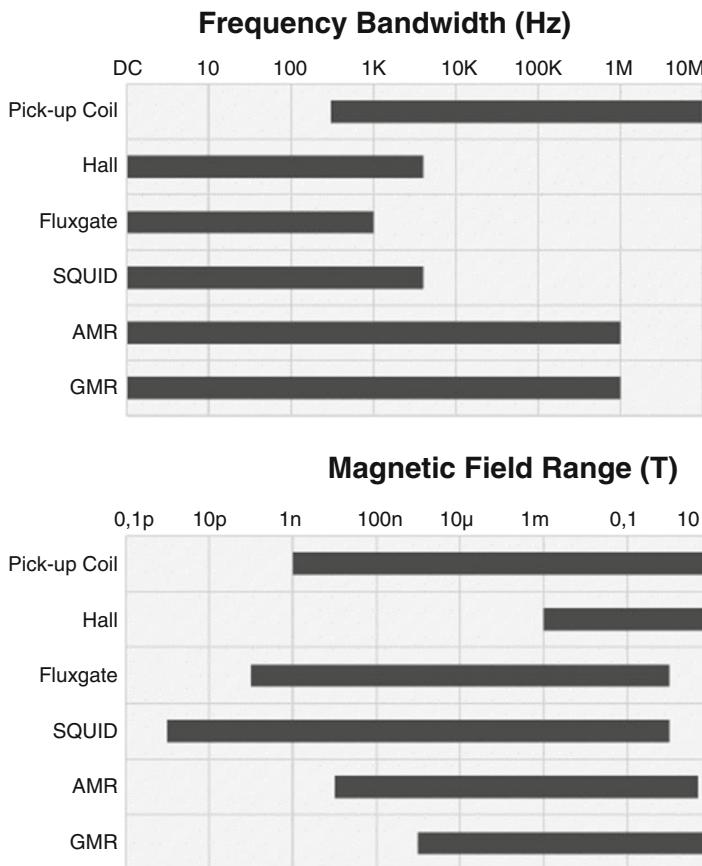
ECT usually involves inspection as an active NDT technique by inducing suitable eddy currents in the specimen and measuring the response due to the material changes. In the measurement process, an ECT probe is mainly used to convert between the electrical signal and electromagnetic energy from the ECT instrument to the inspection materials and structures (Cecco et al. 1986; Harvey 1995; Haus and Melcher 1989). The ECT probe provides a signal perturbation to the instrument, which computes and outputs the inspection information of interest. The measured signal and feature parameters are then analyzed to examine the status of the specimen by using signal processing algorithms online or offline. Therefore, the ECT probe is one of the most important parts of an ECT. In addition, there are many different types of the ECT probe. Unlike other NDT methods, the ECT probe is not a general device, but can be specially designed and utilized according to different purposes and applications.

In Fig. 12, an overview of bandwidth and sensitivity comparison of various magnetic field sensors are illustrated (Ramos and Ribeiro 2014; Jogscies et al. 2015; Davis 1996). ECT probes are usually made as induction coils in various types with complex shapes. On the other hand, magnetic field sensors are sometimes employed to sense low-intensity perturbation electromagnetic fields. Some are very sensitive and able to measure magnetic fields in the region of interest, some are tiny allowing a good spatial resolution, and some present a large bandwidth from DC to the MHz scale. The SQUID is the most sensitive low-field sensor when using low-frequency signals. However, their high cost and difficulty in handling limit their applications in ECT. Fortunately, thin film technology made possible the development of small solid-state magnetic field sensors. Anisotropic magnetoresistance (AMR), giant magnetoresistance (GMR), and tunnel magnetoresistance (TMR) are such sensors which use the magnetoresistance of the ferromagnetic thin films. These sensors have high sensitivity that is almost constant along their wide frequency bandwidth, which makes them good candidates as pickup units in ECT probes for detection of surface and subsurface flaws.

## Types of Sensing Elements for ECT Probes

### Induction Coils

Induction coil-type ECT probes are widely used as they can be designed and constructed easily and can fulfill the inspection requirements in a wide range of applications (Tumanski 2007; Janousek et al. 2009; Martin et al. 2011). The induction coil probe is based on Faraday's induction law that states the voltage induced in



**Fig. 12** Bandwidth and sensitivity comparison between various magnetic field sensors

a coil is proportional to the time derivative of the magnetic flux across it. The sensitivity of induction coils depends on the number of turns, section diameter, permeability of the core, and applied frequency. Generally, the induction coil is used as both excitation and pickup element. Impedance changes caused by the inducing voltage in the induction coils are the main variable to evaluate the properties of the specimen under test. There are two basic probes configurations concerning the coupling between the excitation and the detection coils, i.e.:

- Self-induction probes
- Mutual induction probes

In the case of self-induction probes, the excitation and the pickup coils are identical, and the impedance of the coils is taken as the output signal. In the case of mutual induction probes, the excitation and the pickup coils are separated, and the

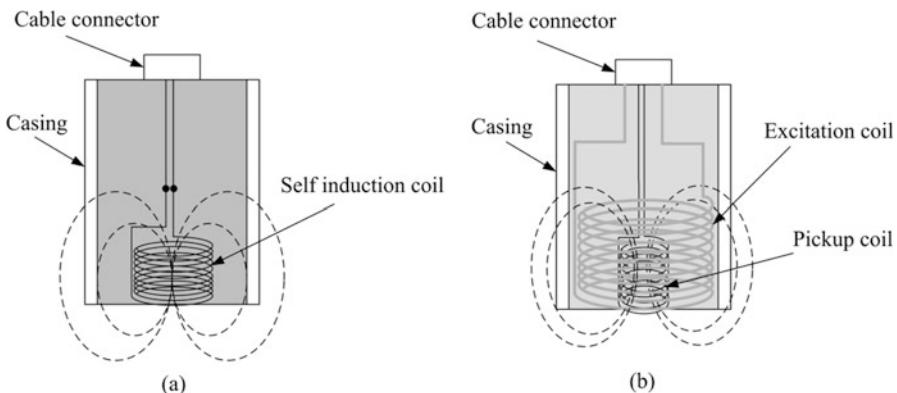
induced voltage in the detection coil is taken as the output signal for defect evaluation. Both probe configurations are shown in Fig. 13.

According to the number and connection between detection coils, they can be mainly operated in the following modes, i.e.:

- Absolute mode, when an ECT probe consists of only one detection coil.
- Differential mode, when an ECT probe consists of two detection coils and they are connected in a magnetically opposite configuration.
- Additive mode, when an ECT probe consists of two detection coils and they are connected in series.
- Mixed mode, when an ECT consists of more than two detection coils.

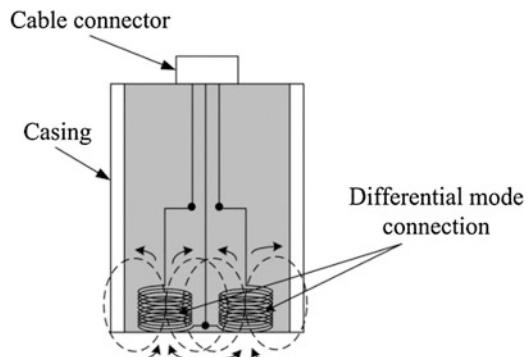
The differential connection of detecting coils, shown in Fig. 14, is the most commonly utilized mode in order to reduce noise due to probe wobbling.

The induction coils present some advantages due to their robustness, ease of manufacture, and shape/size to suit the inspection task and are particularly suitable



**Fig. 13** Basic types of ECT probes: (a) self-induction probe and (b) mutual induction probe

**Fig. 14** Differential connection of pickup coils



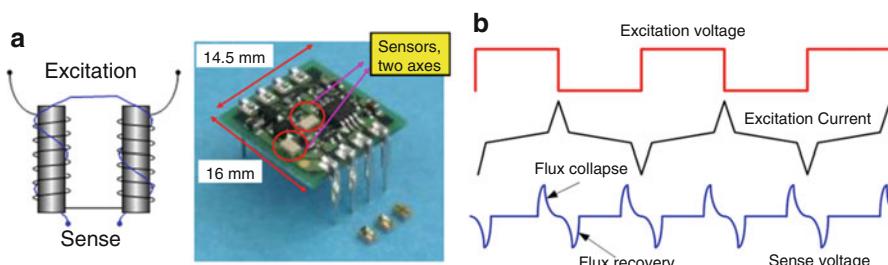
for scanning surfaces. The main drawback of the induction coils is the necessity of using higher frequencies, because the induced voltages are proportional to the amplitude of the magnetic flux and the frequency.

### Fluxgate Sensors (Ripka 2003; Xie 2012)

Fluxgate sensor is one kind of magnetic sensor for picking up the magnetic field directly. The name fluxgate clearly derives from the action of the core gating flux in and out of the sensing coil. The fluxgate sensor consists of a ferromagnetic material wound with two coils, a driving and a sensing coil as shown in Fig. 15a. The measuring principle of typical fluxgate sensor is based on the second harmonic principle. This process is shown in Fig. 15b as idealized waveforms, and it can be seen that the sensing voltage is twice the frequency of the excitation. When a sufficiently large sinusoidal current is applied to the drive coil, the core reaches its saturation magnetization once each half-cycle. If an external magnetic field is applied to the sensor, the symmetry of the cycling saturation is disturbed, and a second harmonic amplitude of the reference signal corresponding to the external field strength appears. For readout, the second harmonic is extracted and rectified. The voltage associated with this harmonic is proportional to the external magnetic field. Fluxgate sensors can provide field sensitivities in the range from pT to  $\mu$ T. The frequency response of the sensor is limited by the excitation field and the response time of the ferromagnetic material. The upper limit on the frequency is a few tens of kHz.

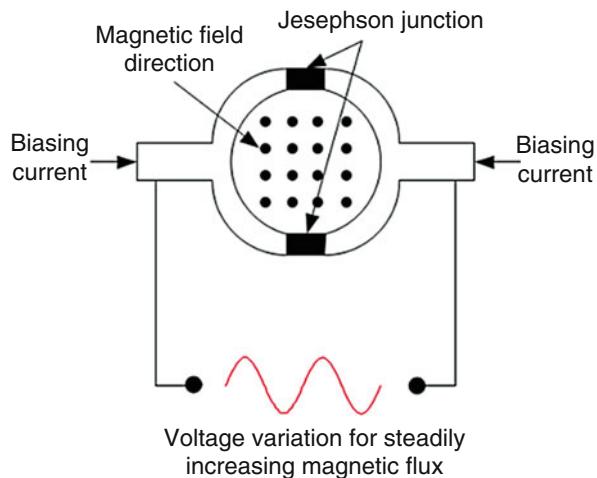
### Superconductivity Quantum Interference Device (SQUID)

The most sensitive sensors for magnetic fields are SQUIDs (Cochran and Carr 1995; Kreutzbruck and Krause 2002). The basic phenomena governing the operation of SQUID devices are flux quantization in superconducting loops and the Josephson effect. The Josephson effect refers to the ability of two weakly coupled superconductors to sustain at zero voltage a supercurrent whose magnitude depends on the phase difference between the two superconductors. When the current passed through a Josephson weak link exceeds the maximum current that it can support, a voltage appears across it. Applying a small magnetic field parallel to the plane of the junction, the phase difference between the two superconducting electrodes varies



**Fig. 15** (a) Fluxgate sensor and circuit board, (b) output waveform of fluxgate sensor

**Fig. 16** Working principle of SQUIDs

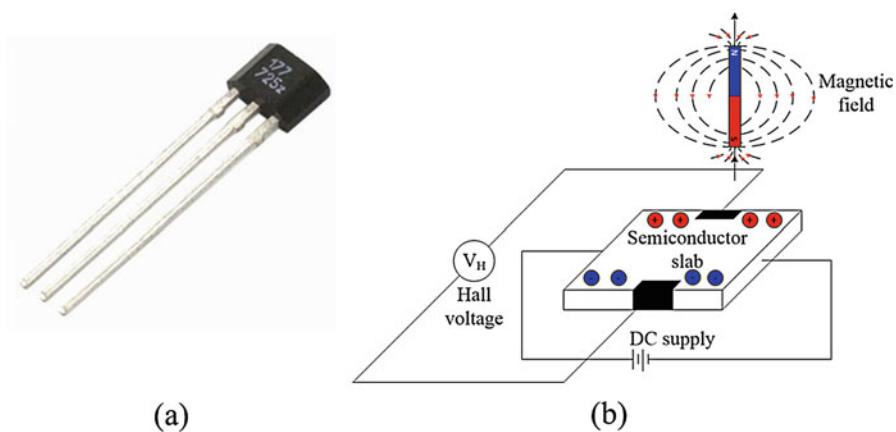


linearly along a direction transverse to the direction of the magnetic field as shown in Fig. 16. This variation of phase difference causes supercurrents tunneling across the different area elements of the junction to interfere and results in a variation of the maximum current of the Josephson junction with the applied magnetic field.

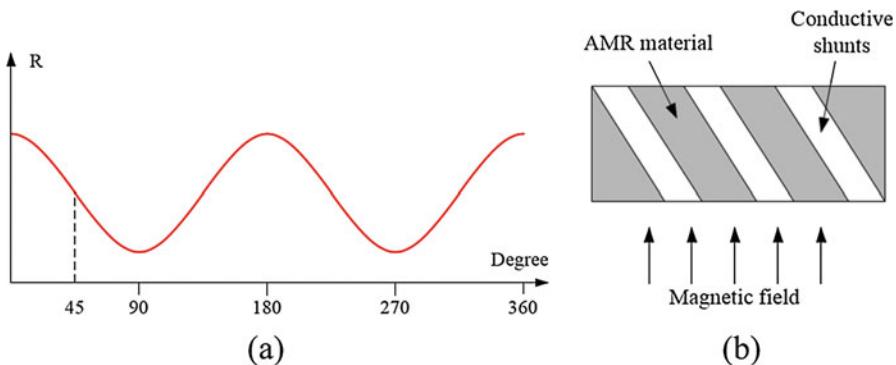
SQUID systems for NDE have high sensitivity ( $\approx 10 - 100 \text{ fT/Hz}^{1/2}$ ), wide bandwidth (from DC to 10 kHz), and broad dynamic range ( $> 80 \text{ dB}$ ). Nevertheless, their unavoidable high cost and difficult handling constitute an obstacle to the wide use of SQUID technology in NDT. Another limitation results from the difficulty in taking advantage of the high sensitivity of SQUIDs due to the presence of noise in the ECT inspections.

### Hall Sensors

Hall sensors are based on the Hall effect (Kosmas et al. 2005; Hashimoto et al. 1993; He et al. 2010), i.e., a magnetic field generates a voltage in a Hall element. The element consists of a thin piece of a semiconductor material with a continuous current passing through it. When the device as shown in Fig. 17a is placed within a magnetic field, the magnetic flux exert a force on the semiconductor material which deflects the charge carriers and give rise to an output voltage called Hall voltage proportional to the magnetic flux density ( $B$ ) as shown in Fig. 17b. The effect occurs either with direct or alternating fields and the voltage across the two parallel faces varies at the same frequency as  $B$ . The Hall effect is very small in metallic conductors but may be significant in semiconductors. Inexpensive Hall sensors are generally made of silicon. More sensitive sensors can be made of semiconductors that have higher electron mobility than silicon. Hall sensors can measure either a constant or a varying magnetic field. The magnetic flux density can be measured from  $\mu\text{T}$  to  $\text{T}$  with a resolution better than 100 nT, but it has a limitation on the frequency. They can be operated over a wide temperature range limited only by packaging and lead attachment to the semiconductor.



**Fig. 17** (a) Typical Hall sensor, (b) principle of the Hall effect-induced output voltage



**Fig. 18** (a) Variation of resistance as a function of angle between the current and magnetization, (b) structure of conductive shunts

### Anisotropic Magnetoresistance Sensors

The AMR effect occurs in ferromagnetic materials (Ripka et al. 2003; Jander et al. 2005; He et al. 2011). It is a property of ferromagnetic materials in which there is a dependence of electrical resistance on the angle between the direction of the electric current and direction of magnetization. The effect arises from the simultaneous action of magnetization and spin-orbit interaction, and its detailed mechanism depends on the material. The AMR sensor is termed anisotropic because it is associated with the angle between the electric current and the direction of magnetization. The AMR effect is described by a variation in the resistance due to a change in the scattering cross section of atomic orbitals distorted by the magnetic field. The resistance produced by scattering is given as a function of the angle. As illustrated in Fig. 18a, the resistance is maximum when the direction of magnetization is parallel to the direction of current and minimum when it is perpendicular to the current. Maximum sensitivity and linearity occur at 45°. This

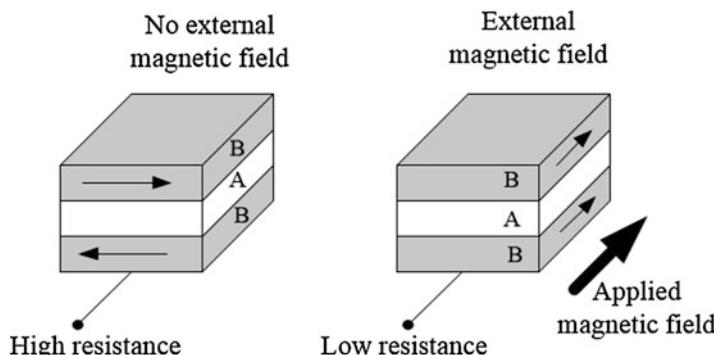
alignment of magnetization and current directions is commonly achieved by patterning diagonal strips of highly conductivity metal onto the more resistive AMR material. The magnetization remains preferentially along the direction of the device, and the current will then run perpendicular to these diagonal stripes. When an external magnetic field is applied, the magnetization vector rotates with a resulting change in resistance.

### Giant Magnetoresistive Sensors

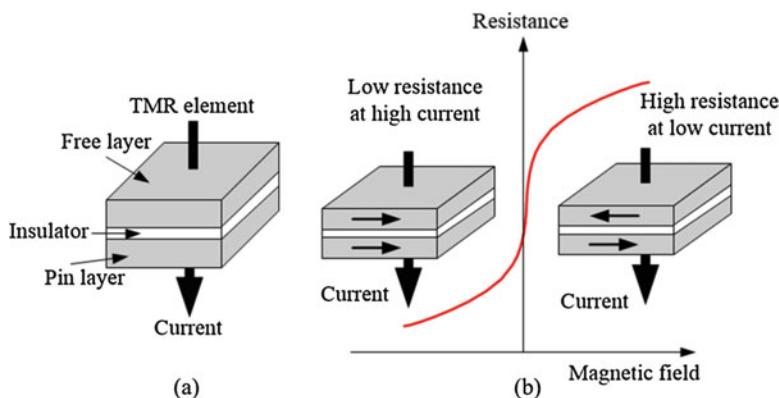
The principle of the GMR sensor is based on the quantum mechanical magnetoresistance effect observed in multilayers composed of alternating ferromagnetic and nonmagnetic conductive layers (Jogschies et al. 2015; Jander et al. 2005; Kim et al. 2010; Postolache et al. 2012). A GMR sensor consists of magnetic metallic multilayers such as Fe/Cr, Co/Cu, and NiCo/Cu in which ferromagnetic layers are separated by nonmagnetic layers a few nanometers thick. The term "giant" in GMR refers to the large change in resistance, which decreases typically 4% to over 20%, when the devices are subjected to a magnetic field. As shown in Fig. 19, applying a magnetic field to these metallic multilayers results in a significant reduction of the electrical resistance of the multilayer due to the spin-dependent scattering of electrons at the interfaces. In a bulk material form, conduction electrons in these materials can travel a distance before "scattering" changing spin information, or changing direction, due to collision with other atomic particles. However, in materials that are very thin, an electron cannot travel the maximum mean free path length, it is more likely that the electron will reach the boundary of the material and scatter there, rather than scatter off another atomic particle. This results in a difficulty for conduction electrons to travel in very thin materials and a higher electrical resistivity. In order to take advantage of this effect, GMR films are manufactured with very thin layers of alternating magnetic and nonmagnetic materials.

### Tunnel Magnetoresistance Sensors

The TMR sensor is almost the same as the GMR element in principle (Jander et al. 2005; Pelkner et al. 2015; Li et al. 2016). However, the current flows perpendicular



**Fig. 19** The GMR phenomenon for detection of magnetic flux density



**Fig. 20** (a) The basic structure elements of TMR sensor, (b) principle of TMR

to the film's surface in a TMR element, while it flows parallel to the film surface in a GMR element. The basic structure of the TMR element is shown in Fig. 20. In this structure a barrier layer of thin insulator, 1 to 2 nm thick, is sandwiched between two ferromagnetic layers made using an advanced thin film processing technology. Although the magnetization direction of the pin layer is fixed, the magnetization direction of the free layer changes according to the direction of the external magnetic field. The electrical resistance of the TMR element changes along with this change in the free layer. The electrical resistance becomes smallest when the directions of magnetization of the pin layer and free layer are parallel, causing a large current to flow into the barrier layer. When the magnetization directions are antiparallel, the resistance becomes extremely large, and almost no current flows into the barrier layer.

The output of a TMR sensor is about 20 times higher than that of an AMR sensor and 6 times higher than that of a GMR sensor.

### Design of ECT Probes Based on Numerical Analysis

Conventionally, ECT probes have been designed on the basis of empirical strategies. The most common type of ECT probe is usually an induction coil. In design of ECT probes, optimization for high sensitivity to expected flaws, and possibility to distinguish flaw variables such as length and depth of expected flaws are the main requirements. Therefore, the excitation coil should induce eddy currents with high current density and such distribution that eddy current lines are significantly perturbed when flaws are present. The pickup coil should be such that the perturbation field is sensed with maximum output. There are three basic techniques of probe design, which can be classified as follows: (1) experimental or empirical design, (2) analytical design, and (3) numerical design. A practical way to design a probe would be to start with analytical expressions (exact or approximate), design a probe

based on some set of initial requirements, construct the probe, and then evaluate its performance experimentally. If necessary, the process can be repeated until an acceptable design is obtained.

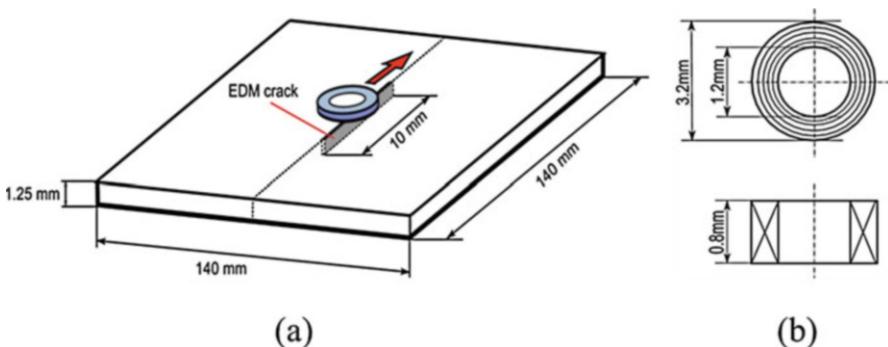
Recently, the development of electromagnetic field computation techniques (Takagi et al. 1994; Reis et al. 2002; Takagi et al. 1998) enabled the application of numerical methods to the detectability analysis of ECT probes. The numerical design of probes has several advantages, such as:

- The probe, with all its components (coils, core, and shield) and the surrounding medium, can be analyzed. the probe characteristics in the actual test environment can be obtained.
- A more accurate design is obtained before the probe can be actually built by numerically experimenting with the probe parameters.
- The numerical technique is applicable to situations that cannot be analyzed analytically or experimentally (subsurface discontinuities, layered materials, and others).

The following optimal design in this section describes numerical approaches based on section “[Numerical Methods for Eddy Current Testing](#)” involving an iterative approach in which the test results from a specific design lead to the optimal design of ECT probes.

### Absolute Pancake Probe for Surface Crack Testing

The JSAEM benchmark problem of step 2 for design of a simple pancake-type ECT probe is employed for numerical analysis of the eddy current signals based on  $\mathcal{A} - \phi$  and Ar methods (JSAEM 1997). The JSAEM benchmark problem step 2 as shown in Fig. 21a needs to predict the impedance signal from the ECT probe when it scans past an EDM notch with dimensions of 0.2 mm width, 10 mm length, and 50% depth, which is set in the center of a 1.25-mm-thick conductor plate. The given conductivity of the plate is 1 MS/m to represent Inconel steel used in steam generator tubes in PWR nuclear power plants. The ECT probe is a pancake-type coil with a thickness of 0.8 mm, an outer diameter of 3.2 mm, and an inner diameter of 1.2 mm as shown in



**Fig. 21** Model of the JSAEM benchmark problem step 2: (a) target plate, (b) probe coil

Fig. 21b. The number of turns in the coil is 140. The probe lift-off distance, total current, and excitation frequency are 0.5 mm, 1 A, and 300 kHz, respectively.

According to the setup of the model problem, the aforementioned numerical modeling for three-dimensional eddy current field and prediction of the eddy current detection signal are utilized to verify the validity of the numerical methods. Figure 22 gives a comparison between FEM-BEM code using the  $A - \phi$  formulation, FEM code using the  $A_r$  method, and experimental results. The ECT signals of the real and imaginary parts as scanning curves and in the impedance plane are shown in Fig. 22a and b, respectively. From the results, it can be seen that the computation of the selected numerical methods is in good agreement with the experimental results. Therefore, using numerical modeling for the ECT probe design and prediction of the ECT signal

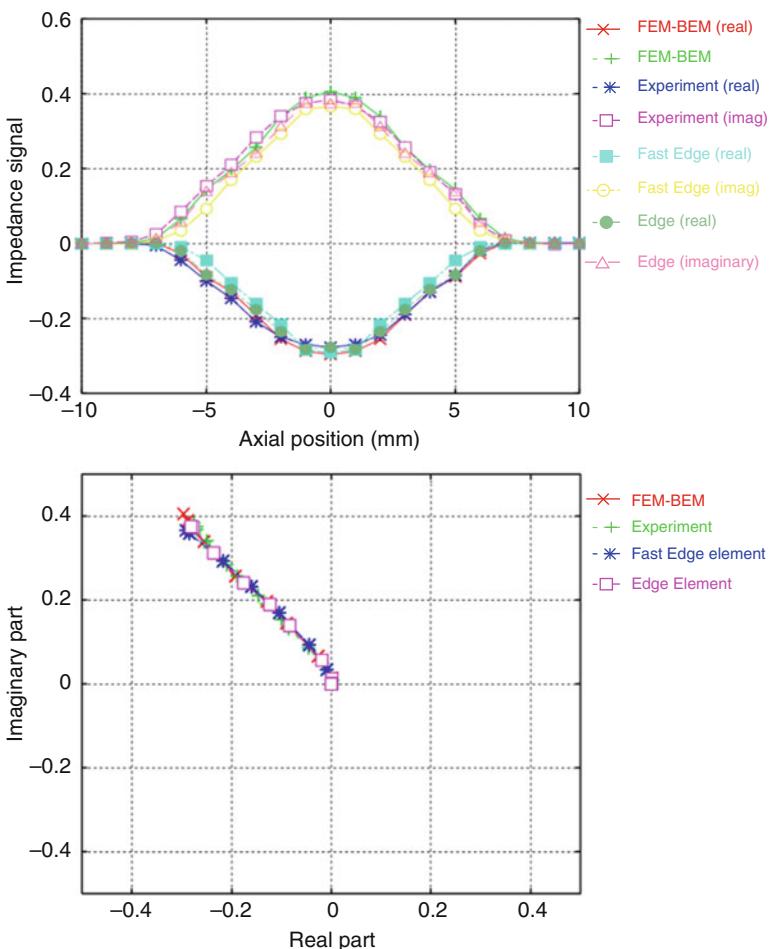
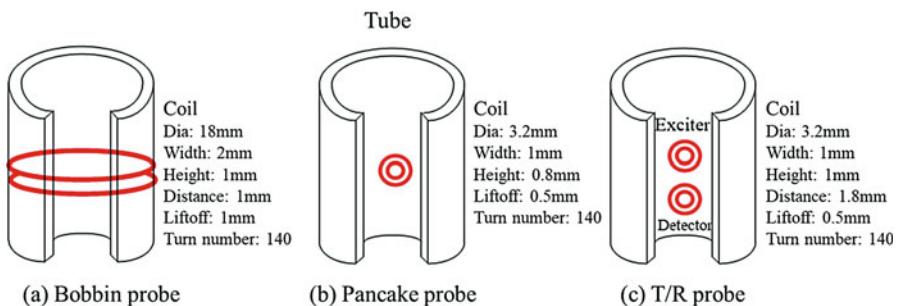


Fig. 22 Comparison of signals from the FEM-BEM,  $A_r$  code, and experimental results. (a) A scope of ECT signal, (b) impedance plane

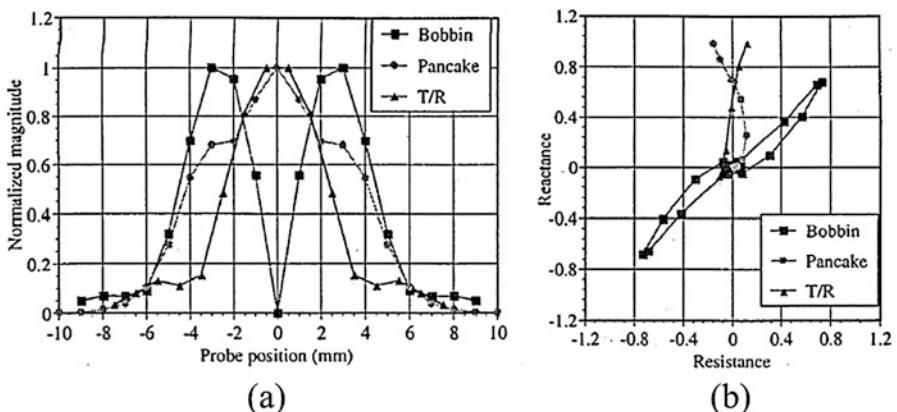
offers certain advantages over other design techniques, in terms of efficiency and resources requirement.

### ECT Probes for Tube Inspection

In this case, ECT applied to the inspection of the steam generator tubes in PWR nuclear power plants is simulated based on the FEM-BEM hybrid code (Takagi et al. 1997; Maeda et al. 1997). The considered specimen and three kinds of probes, i.e., a conventional bobbin-type probe, a pancake-type probe, and a TR-type probe, are shown in Fig. 23. The origin of the probes is defined at the center of the axial crack, and the probe scanning paths run parallel to axis of the tube. The sampling points of the pickup signals as the coil moves parallel to the axial cracks are set to 21, from  $x = -10$  mm to 10 mm. The signals of the three probes are computed and compared in Fig. 24 for a frequency of 150 kHz. The magnitude is normalized by the maximum values of each signal. In the cases of both the pancake-type and TR-type probes, peak values are observed at the position where the centers of the probes overlap those of the crack. The trajectories of the signals are different from each other. This simulation makes



**Fig. 23** Tubes and three kinds of ECT probe



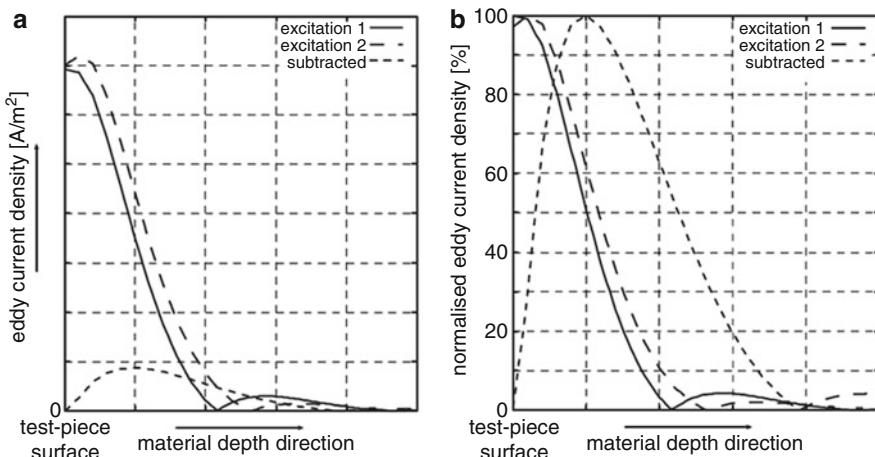
**Fig. 24** Comparison of ECT signals from a bobbin probe, a pancake probe, and a TR probe.  
(a) Time history of magnitude, (b) impedance plane plot

it possible to evaluate the shapes of the signals from cracks. The signals indicating the presence of a crack have various aspects, depending on the probes used. In contrast to the bobbin-type probe, maximum values in the cases of the pancake-type and TR-type probes occur at the origin. The bobbin-type probes were used in the in-service inspection of tubes for a long time. Their scanning speeds are fast. However, they are lacking in resolution into the position of cracks in the circumferential direction of the tubes in comparison to the other two probes. Although the rotating pancake probe was developed to alleviate the problem of resolution, it has a disadvantage in scanning speed. To obtain higher S/N ratio, the TR-type probes were developed for practical applications.

### ECT Probe for Deep Crack Detection

To enhance the detectability of ECT for sizing deep surface-breaking cracks in thick structures, a special ECT probe was developed (Janousek et al. 2005; Chen et al. 2004). The key idea is to suppress eddy currents on the surface of a specimen to realize deeper penetration of eddy currents. Through superposition of several exciting fields, which create similar EC patterns of the same density on surfaces at certain points, different EC distributions along the material depth direction can be realized. These conditions can be accomplished by using coaxial exciting coils that have the same shape and orientation. By adjusting their location with respect to the particular point of interest, as illustrated in Fig. 25, the eddy current density profiles in the depth direction at the point can be controlled. Subtraction of the magnetic fields due to these two excitation coils creates a new eddy current distribution which is marked as “subtracted” in Fig. 25. As can be seen, the eddy current density can be suppressed to zero at the surface but shows a peak inside the material.

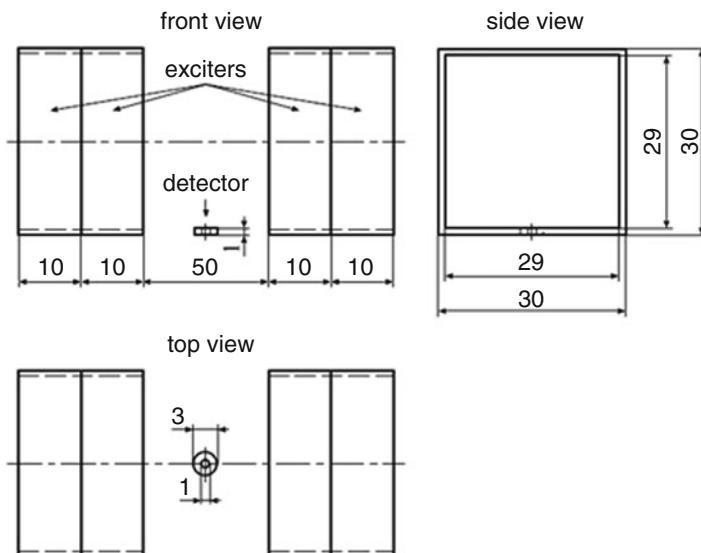
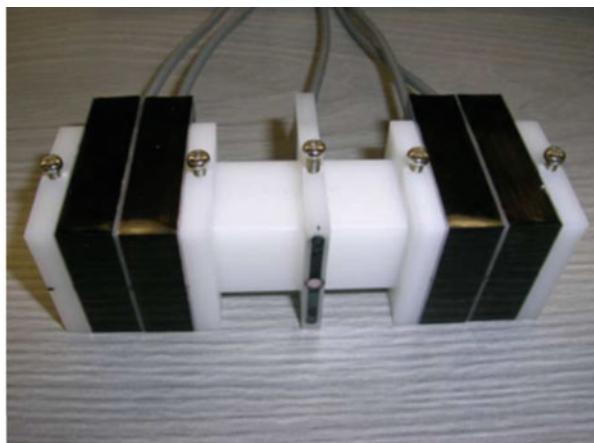
In practice, the probe is of mutual inductance type and consists of four coaxial exciting coils and one detecting coil. The inner exciting coils and the outer ones are connected in series, respectively. The two groups of the exciting coils are driven by



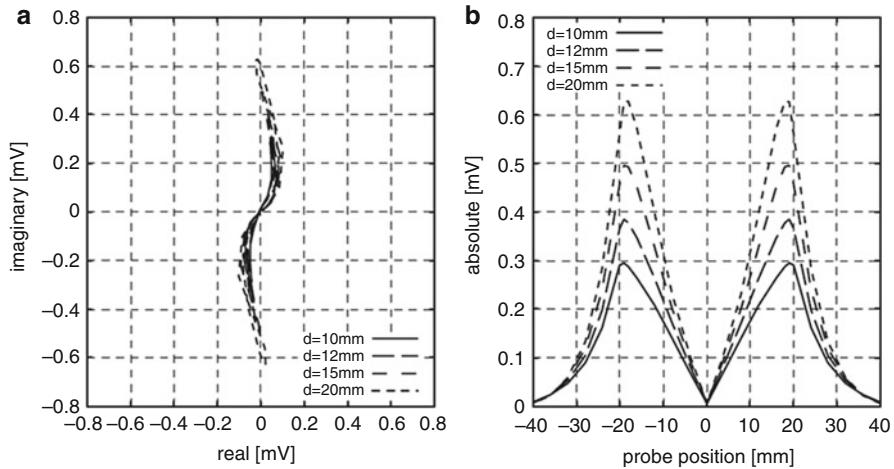
**Fig. 25** Eddy current distribution along material depth. (a) Absolute values, (b) relative values

two harmonic currents with the same frequency but different phases. Proper adjustment of the amplitude and phase of the excitation currents allows one to change the profile of the eddy current density distribution under the detecting coil in such a way that zero density is obtained on the surface of the tested material. In this way, the probe is capable for sizing deep surface cracks. Detailed configuration of the probe is shown in Fig. 26.

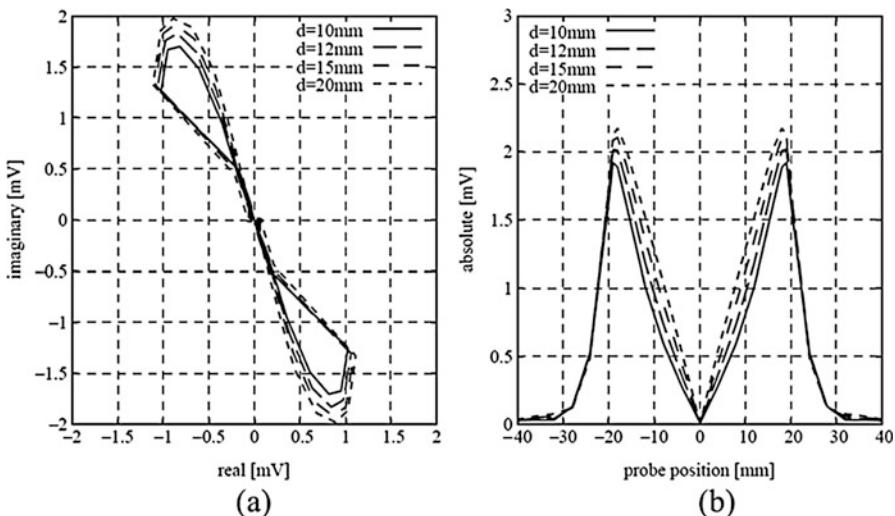
Numerical analysis as well as experimental verifications revealed, as shown in Fig. 27, that the new ECT probe for deep crack detection can be used for evaluation of surface-breaking cracks with a depth up to 20 mm. In contrast, as shown in



**Fig. 26** Configuration of the ECT probe for deep crack detection



**Fig. 27** Simulated crack signals with the ECT probe for deep crack detection. (a) Impedance plane, (b) time history of magnitude



**Fig. 28** Simulated crack signals with the ECT probe with standard excitation. (a) Impedance plane, (b) time history of magnitude

Fig. 28, an ECT probe with conventional excitation can be employed only up to a depth less than 10 mm under the same conditions.

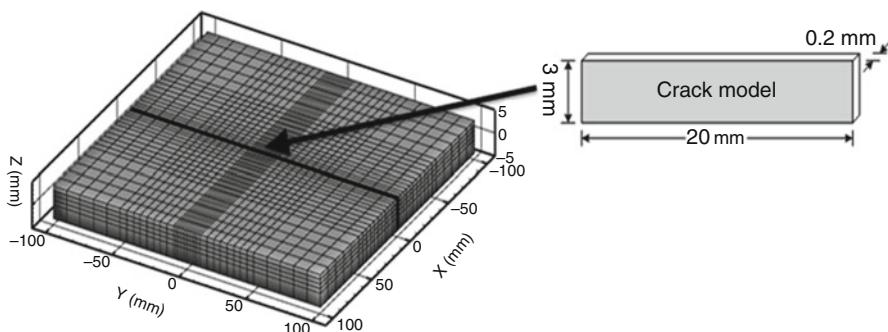
### ECT Probe for Crack Detection Using Magnetic Field Measurement

To enhance the design of ECT for magnetic field measurement using magnetic field sensors such as AMR, GMR, or TMR, numerical methods can be used to simulate the

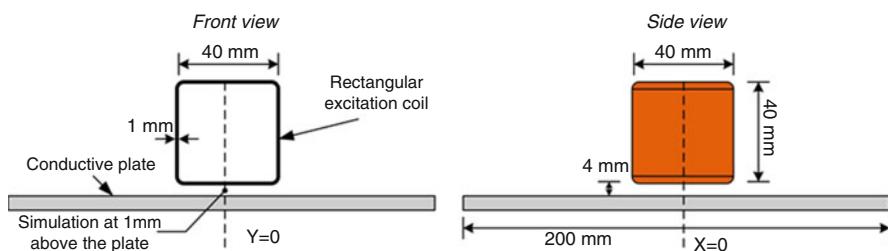
magnetic field perturbation due to a crack in three directions, i.e.,  $x$ ,  $y$ , and  $z$  components. This analysis can help design and construct ECT probes using magnetic field sensors for profiling crack shapes (Jomdecha et al. 2018). A 3D numerical model is developed based on a conducting plate with conductivity of 1.35 MS/m (304 type stainless steel). Plate dimensions were set as  $200 \times 200 \times 5$  mm<sup>3</sup> as shown in Fig. 29. The crack model is in the shape of a notch and is set symmetrically in the center of the conductor plate along the  $y$  direction with its dimensions of width, length, and depth being 0.2 mm, 20 mm, and 3 mm, respectively.

The numerical code using the  $A - \phi$  formulation is utilized to simulate the magnetic flux perturbation due to eddy currents induced by an ECT probe with uniform current excitation. A rectangular excitation coil with dimensions as illustrated in Fig. 30 was utilized to induce the near uniform eddy currents into the conductor plate for analysis of the magnetic flux perturbation nearby the plate. The number of turns of the coil was set to 1000 with a total excitation current of 0.5 A. The excitation coil was discretized into  $10 \times 10$  current filaments of rectangular shape to calculate the source magnetic field. The coil was placed symmetrically at the center of the plate with 4 mm lift-off. The testing frequency was set to 50 kHz.

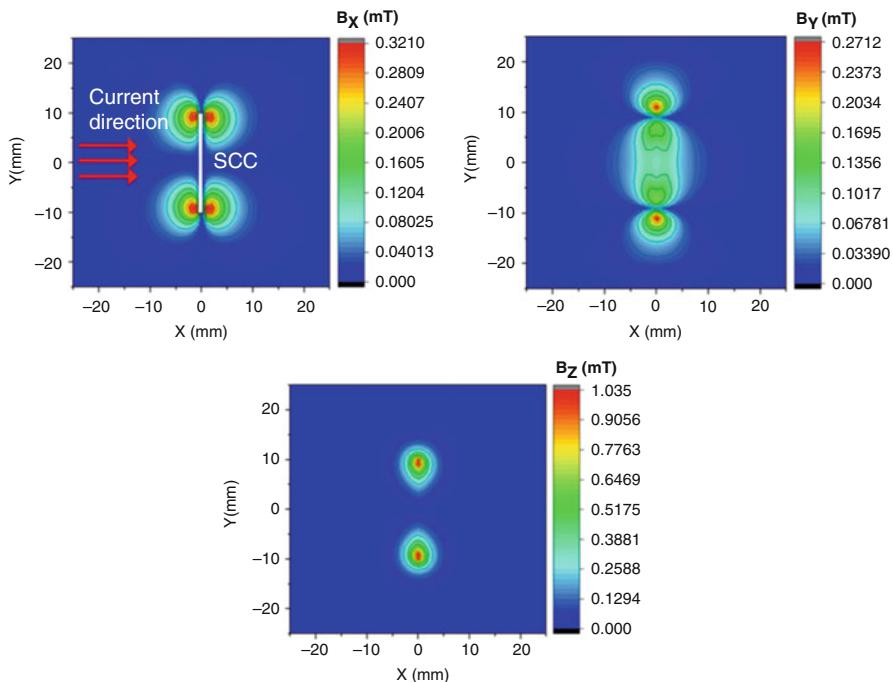
By setting the equivalent crack models at the center of the plate with its length perpendicular to the excitation current, the numerical results for the magnetic flux perturbation due to crack at a plane 1 mm above the conductor plate surface were calculated. Components of the perturbed magnetic field above the crack regions have



**Fig. 29** A model of crack detection using magnetic field measurement



**Fig. 30** Rectangular coil with its dimensions and location used to induce a uniform current



**Fig. 31**  $B_x$ ,  $B_y$ , and  $B_z$  perturbation due to crack perpendicular to the excitation current

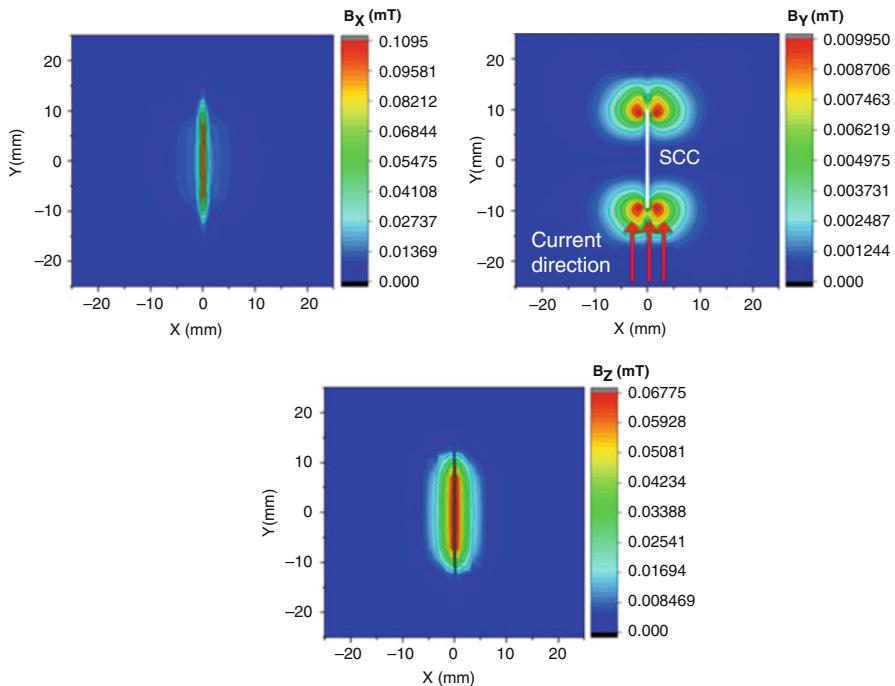
the distributions as shown in Fig. 31. The  $x$  and  $z$  components can be used to clearly indicate the length of the crack. On the other hand, the profile of the  $y$  component distributes throughout the length of the crack region.

As shown in Fig. 32, the numerical results of perturbed distributions of the magnetic field were obtained from the equivalent crack model set laterally to the excitation current. The results show that the  $B_x$  profile can be correlated with the equivalent crack shape. However, the amplitudes of these results are lower than the amplitude of the excitation current perpendicular to the crack direction, i.e., the uniformly induced eddy current produces a lower perturbation in the parallel case.

## Simplified Method for Detectability Evaluation of ECT Probes

### Correlation between the Source Magnetic Field and Induced Eddy Currents

As is well known, the basic difficulty of optimizing the design of ECT probes is caused by the need to determine too many parameters, not only the parameters of a given probe configuration but also the parameters of its shape and arrangement such as the numbers of the excitation and pickup coils which also need to be modified. Therefore, it is difficult to choose the best probe parameters by a normal optimization



**Fig. 32**  $B_x$ ,  $B_y$ , and  $B_z$  perturbation due to a crack parallel to the excitation current

scheme unless the basic structure of a probe and its potential for high performance could be known a priori. In this section, a phenomenological relationship is proposed to describe the relationship between the distribution of the source magnetic flux density and the induced eddy currents for a nonmagnetic conducting plate or tube by summarizing numerical results. To determine this basic structure, some qualitative methods are necessary to predict the crack-probe interaction behavior and the detectability by the ECT probe.

In this section, a phenomenological relationship proposed by the authors is described, which connects the source magnetic field  $\mathbf{B}_0$  of the exciting coil and  $\mathbf{J}_e$ , the eddy current density induced in an unflawed conductor of large curvature radius (plate or tube). Using this  $\mathbf{J}_e - \mathbf{B}_0$  relation and the ring current model for eddy current perturbation due to cracks, which was also proposed by the authors, a simplified approach for the detectability analysis was established. It was demonstrated that the ring current model and the simplified detectability analysis method satisfy the requirements of a qualitative analysis during the probe design. Based on this phenomenological model, a scheme for ECT probe design was developed in (Chen and Miya 1998a and b) by optimizing the pickup arrangement with use of the ring current model and determining the corresponding excitation methods through simplified detectability analyses.

From numerical results of the eddy current distribution for different frequencies and different arrangements of excitation coil, the relation in Eq. (99) was proposed for an unflawed conducting plate and a probe placed over and nearby the center of the plate so it is free of the conductor edge effect.

$$\mathbf{J}_e(x,y,z,t) = \alpha(z,\omega)\mathbf{n} \times \mathbf{B}_0[x,y,z_0(\omega)] \cos [\omega t + \phi(x,y,z)] \quad (99)$$

where  $\mathbf{J}_e(x, y, z, t)$  is the eddy current at a point  $(x, y, z)$  within the conductor at a time  $t$ ,  $\mathbf{n}$  is the normal unit vector of the plate surface at point  $(x, y, z_t)$  with  $z_t$  being the  $z$  coordinate of the top surface (same side as the probe),  $\mathbf{B}_0$  is the source magnetic flux density of the driver coil(s), and  $z_0(\omega)$  is a constant for a given frequency.

In Eq. (99), the coefficients of the magnitude  $\alpha$ , and the phase difference of the eddy current density  $\phi$ , can be approximated by the following formulae referring the analytical solution of the eddy currents in a conducting half space

$$\alpha(z,\omega) = \alpha_0 e^{-\frac{z_t-z}{\delta}} \quad (100)$$

$$\phi = 0.75(z_t - z)/\delta + \frac{\pi}{4}[1 - \cos(d\pi/6R)] \quad (101)$$

where  $\delta = (2/\omega\mu_0\sigma)^{1/2}$  is the skin depth,  $R$  is the outer radius of the driver coil, and  $d$  is the in-plane distance between the field point and the center of the excitation coil. Though  $\alpha_0$  in Eq. (100) is an unknown constant, it will be canceled during the calculation of S/N ratio which is necessary for probe detectability evaluation. The constants 0.75 and 6 in Eq. (101) were calibrated using the numerical results of a pancake coil probe and were verified to be suitable for other probes by comparing their numerical results in the conductor region near the center point of the excitation coil and for the frequencies usually used in ECT inspection of steam generator tubes. As the pickup coils are usually arranged around the excitation coil, Eq. (101) is especially suitable for the calculation of the pickup signals of conventional ECT probes.

Considering a tube of inner radius  $r_{in}$ , the simplified relation in Eq. (99) in a cylindrical coordinate system shown in Fig. 33b becomes the following expression:

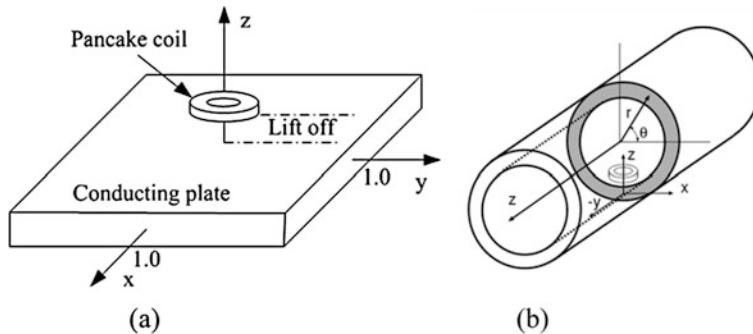
$$\mathbf{J}_e(r,\theta,z,t) = \alpha(r,\omega)\mathbf{n} \times \mathbf{B}_0[r_0(\omega),\theta,z] \cos [\omega t + \phi(r,\theta,z)] \quad (102)$$

while the magnitude coefficient and the phase difference become

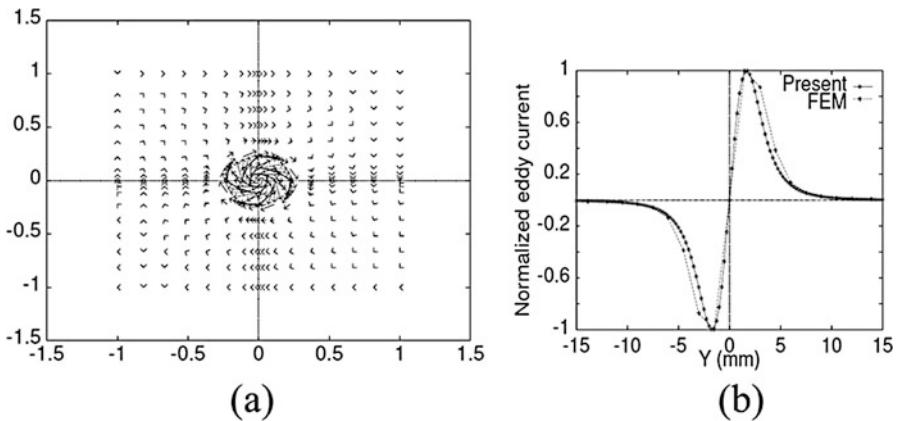
$$\alpha(r,\omega) = \alpha_0 e^{-\frac{r-r_{in}}{\delta}} \quad (103)$$

$$\phi = 0.75(r - r_{in})/\delta + \frac{\pi}{4}[1 - \cos(d\pi/6R)]. \quad (104)$$

To demonstrate the validity of the phenomenological correlations, prediction results of the eddy current induced by a pancake coil parallel to the conducting plate are shown in Fig. 34 for an excitation frequency of 400 kHz. The real part of the eddy current distribution in the plate is shown in Fig. 34a. Comparisons of the eddy



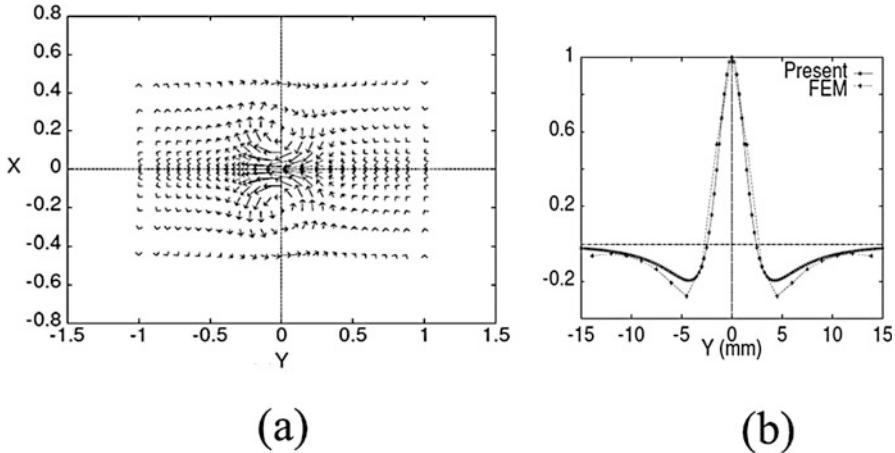
**Fig. 33** Problem definition and coordinate systems, (a) plate geometry, (b) tube geometry



**Fig. 34** (a) Distribution of eddy currents induced by a pancake coil parallel to the conducting plate and (b) comparison of normalized eddy current densities of the simplified  $J_e - B_0$  relation and the FEM-BEM method

current distributions calculated by the FEM-BEM code and by using Eq. (99) are shown in Fig. 34b, which are the eddy current distribution along the line  $y = 0$  at surface with their magnitudes normalized by the maximum values. From the numerical results, it was also found that the eddy currents in different layers have nearly the same distribution, with no significant difference between the distributions of the real and imaginary parts. On the other hand, Fig. 35 shows comparisons of the eddy current distributions calculated by the FEM-BEM code and by using Eq. (102) at the outside layer for a conducting tube by using the excitation of a square coil perpendicular to the tube wall surface. From these results, it was shown that the simplified  $J_e - B_0$  relation is also valid for the tube geometry.

From Eqs. (99), (100), (101), (102), (103), and (104), we conclude that the eddy current distribution at an arbitrary layer of a tube/plate, i.e.,  $r = \text{const}$ , can be described by the distribution of the source magnetic flux density at a given layer



**Fig. 35** (a) Distribution of eddy currents induced by a square coil perpendicular at the tube surface and (b) comparison of normalized eddy current densities of the simplified  $\mathbf{J}_e - \mathbf{B}_0$  relation and the FEM-BEM method

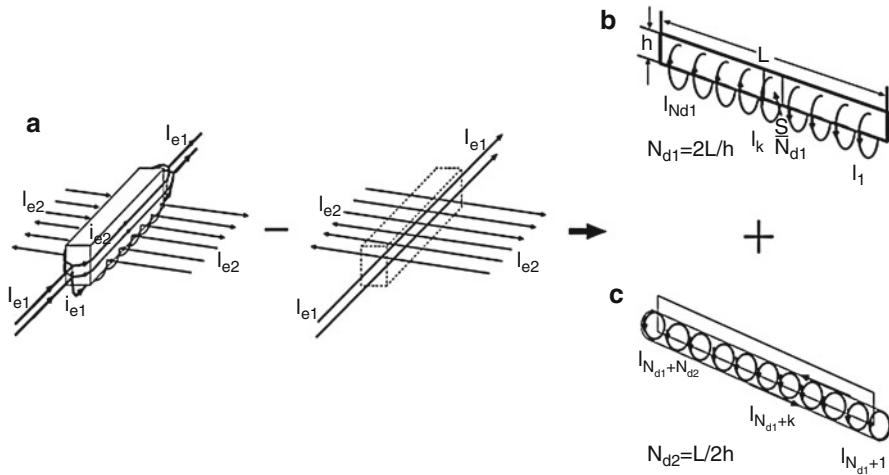
with  $r_0(\omega)$  or  $z_0(\omega)$  which only depends on the excitation circular frequency  $\omega$ . In other words, the eddy current in a nonmagnetic tube/plate conductor can be predicted from the source magnetic field directly by using these formulae for excitation coils satisfying the conditions that it is located far from the edge of the conductor and excited at the frequency of normal ECT inspection. In Eqs. (100) and (101), we assumed that the magnitude of eddy currents decays exponentially according to the skin effect and the phase difference changes linearly in the direction of the conductor thickness. These assumptions are used in the analytical solution of the eddy currents in a conducting half space in a uniform magnetic field (Chen et al. 2004), i.e.:

$$\mathbf{J} = \mathbf{J}_0 e^{-z(1+j)\sqrt{\omega\mu\sigma/2}} e^{j\omega t} \quad (105)$$

as well as in many FEM numerical results for the eddy current distribution under various conditions.

### Modeling of the Perturbed Eddy Current Due to a Crack

To consider the influence of the crack width on the eddy current perturbation, a simplified model was proposed in Chen et al. (1997) by introducing an additional group of eddy current loops and simplifying the current loops as current rings as shown in Fig. 36 (hereafter we call it the ring current model). The ring current set shown in Fig. 36b is used for the approximation of the eddy current perturbation for the component perpendicular to the crack's plane, while the set shown in Fig. 36c is the perturbed eddy current under the crack  $i_{e1}$  for the eddy current component parallel to the crack's plane. The effect of the crack-parallel component, which flows on the two sides of the crack plane, is neglected in the model as the corresponding current loops appear in pairs and in opposite directions.



**Fig. 36** Concept diagram of the ring current model, (a) eddy current flows in the conductors with and without crack, (b) ring current set for the perpendicular eddy component, (c) ring current set for the parallel eddy current component

In the case of a shallow crack, it is reasonable to assume that the centers of the ring currents are at the bottom line of the crack and their radius is the crack depth. The magnitudes of the ring current,  $I_k$ , can be approximated by the corresponding component of the unflawed eddy current at the center of the  $k$ -th current ring. In addition, a coefficient  $\beta$  is introduced in order that the difference between the ring current and the actual path of the perturbed eddy current be considered. As the ring current set (c) is equivalent to the eddy current component flowing under the crack bottom, a coefficient  $\alpha_1 = i_{e1}/I_{e1}$  is also introduced as the proportion of  $i_{e1}$  in the total parallel component  $I_{e1}$  as shown in Fig. 36a. It was verified that these coefficients can be considered approximately as constants for different probes if the crack is unchanged. From the above considerations, the ring current  $I_k$  can be calculated from the source magnetic flux density  $\mathbf{B}_0$  by using Eq. (99) or (102) and the modification coefficients as

(i) For the perpendicular ring currents,

$$I_k = \beta \frac{S_1}{N_{d1}} \mathbf{J}_e \cdot \mathbf{n}_{ck} = \beta \frac{S_1}{N_{d1}} \alpha_0 e^{-\frac{r-r_m}{\delta}} \mathbf{n} \times \mathbf{B}_0[r_0(\omega), \theta_k, z_k] e^{j\phi_k} \cdot \mathbf{n}_{ck}, \quad (106)$$

$$k = 1, 2, \dots, N_{d1},$$

(ii) For the parallel ring currents,

$$I_k = \beta \alpha_1 S_2 \mathbf{J}_e \cdot \mathbf{t}_{ck} = \beta \alpha_1 S_2 \alpha_0 e^{-\frac{r-r_m}{\delta}} \mathbf{n} \times \mathbf{B}_0[r_0(\omega), \theta_k, z_k] e^{j\phi_k} \cdot \mathbf{t}_{ck} \quad (107)$$

$$k = N_{d1} + 1, N_{d1} + 2, \dots, N_{d1} + N_{d2}$$

where  $\theta_k$  and  $z_k$  are the coordinates of the center of the  $k$ -th ring current,  $\phi_k$  is the initial phase at the center of the ring current,  $S_1$  is the area of the crack plane, and  $S_2$  is the cross sectional area of the crack. In addition,  $\mathbf{n}_{ck}$  and  $\mathbf{t}_{ck}$  are, respectively, the unit vectors normal to the crack plane, and along the crack length,  $N_{d1}$  and  $N_{d2}$  are the numbers of the current rings.

### **Scheme for ECT Probe Evaluation Based on the Phenomenological Models**

If we consider the signal-to-noise ratio (S/N ratio) as the major measurement of the probe performance, the probe evaluation then becomes a problem to predict the S/N ratios under selected conditions. Given that the signals from the lift-off change and/or the probe inclination are the major noise sources, which are difficult to eliminate using the usual filtering technique, we choose the S/N ratio against these noise as the quantitative measures of a probe. Therefore, we need to develop a simplified method for S/N ratio calculation of new probes. In practice, this can be done by using the ring current model for crack signal and the phenomenological relationship in Eq. (99) for noise signals directly.

### **The Impedance Change Due to a Crack**

The impedance change due to a crack can be considered as the variation of the pickup signals caused by the corresponding ring currents, i.e., the impedance change of a pickup coil of  $N$  turns is as follows (Takagi et al. 1994):

$$\Delta Z = -\frac{N}{I_0} j\omega \sum_p \oint_{\Gamma_p} \mathbf{A}_{\text{ring}} \cdot dI_p \quad (108)$$

with the vector potential  $\mathbf{A}_{\text{ring}}$  of the ring currents calculated using the Biot-Savart's law as

$$\mathbf{A}_{\text{ring}} = \frac{\mu_0}{4\pi} \sum_{k=1}^{N_{d1}+N_{d2}} \int_{\Gamma_{\text{ring},k}} \frac{\mathbf{I}_k}{\bar{r}} dl_k \quad (109)$$

Combining Eqs. (108) and (109), the impedance change due to a crack can be formulated as.

$$\Delta Z = -\frac{\mu_0}{4\pi I_0} N j\omega \sum_{p=1}^N \oint_{\Gamma_p} \sum_{k=1}^{N_{d1}+N_{d2}} \int_{\Gamma_{\text{ring},k}} \frac{\mathbf{I}_k}{\bar{r}} dI_k \cdot dI_p \quad (110)$$

In these equations,  $N_{d1}$  is the number of the ring currents of set (b) which was chosen as the number shown in Fig. 36,  $N_{d2}$  is the number of the ring currents of set (c),  $I_k$  is the magnitude of the  $k$ -th equivalent ring current given by Eqs. (106) and (107),  $\bar{r}$  is the distance between the source and field points,  $I_0$  is the virtual current in the pickup coil which transform the EMF into impedance, and  $N$  is the number of

turns of the pickup coil. The integral route  $\Gamma_p$  corresponds to the  $p$ -th turn of the pickup coil, and  $\Gamma_{\text{ring},k}$  is the route of the  $k$ -th ring current.

### Impedance Signal from an Unflawed Tube

Once the eddy current has been predicted, the impedance of the pickup coils can be easily calculated from the predicted eddy current by using Eq. (108) again with the vector potential of eddy current  $A_{\text{ring}}$  replaced by  $A_e$  and calculating it with Biot-Savart's law

$$A_e(r) = \frac{\mu_0}{4\pi} \int_V \frac{\mathbf{J}_e(r')}{|r - r'|} dV' \quad (111)$$

and using  $\mathbf{J}_e$  as that given in Eqs. (102), (103), and (104). Finally, Eq. (112) gives an explicit formula connecting the source magnetic field to the pickup impedance signal as

$$Z = -\frac{\mu_0}{4\pi} \frac{\alpha_0 N}{I_0} j\omega \sum_{p=1}^N \oint_{\Gamma_p} \int_V e^{-\frac{|r'-r_{in}|}{\delta}} \mathbf{n} \times \mathbf{B}_0[r_0(\omega), \theta', z'] \frac{e^{j\phi(r', \theta', z')}}{|r - r'|} dV' \cdot dI_p \quad (112)$$

where  $V$  is the volume of the whole conductor and  $r$  and  $r'$  are the position vectors of the field and source point.

Using Eq. (112), the noise signals due to an unexpected lift-off change or probe inclination can be predicted by subtracting the standard signal  $S_{\text{std}}$  (signal at a standard lift-off and zero probe inclination) from the disturbed ones  $S_{\text{std+sway}}$  (signal with lift-off change or/and probe inclination). Using these noise signals and the crack signal  $S_{\text{ring}}$  of Eq. (110), the S/N ratio can be finally obtained by using the following definition:

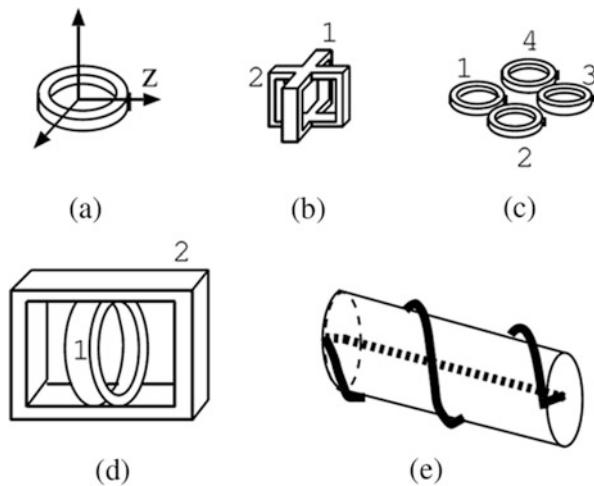
$$S/N = S_{\text{ring}} / (S_{\text{std+sway}} - S_{\text{std}}) \quad (113)$$

Comparing with the conventional methods, only very modest computational effort is necessary for the S/N ratio calculation with use of the approach described above. Though the absolute accuracy of this simplified method is limited, it usually can give satisfactory answers for ranking the detectability of different probes.

### Optimal Design of ECT Probes for Crack Detection

For optimal design of ECT probes sensitive to a crack in tube, it is fundamental that the eddy currents induced by the probe contain large components in both axial and circumferential directions. The maximum eddy current perturbation usually occurs, when the excitation coil is located above the end or the center of a crack. Examples of typical excitation coil structures of an ECT probe which can induce such an eddy

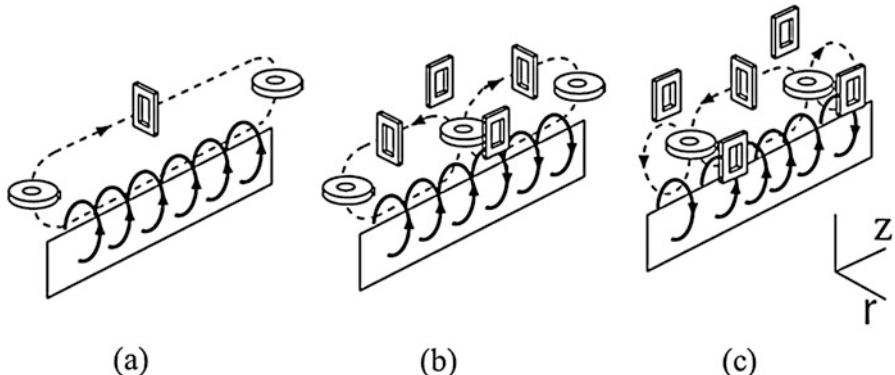
**Fig. 37** Examples of possible design of excitation coils



current distribution are shown in Fig. 37, including pancake coil, plus-point coils, four probe coils, hybrid coils, and spiral coils, respectively.

To optimize the locations to place the pickup coil, the basic configuration of a probe must be determined in the probe design. Referring to Kurokawa (1997) and Popa et al. (1997), differential-type probes with mutual inductance are relatively robust against lift-off noise. It is also shown that the noise due to probe inclination can be significantly reduced by summing the signals of axial pickup coils as well as those of circumferential ones separately. In order to reduce both the lift-off and inclination noise, i.e., wobbling noise, a possible approach is to arrange the pickup coils symmetrically about the center of the excitation coil in both the axial and circumferential directions. In this case, the lift-off and inclination noise can be reduced simultaneously by subtracting the sum signals in the axial direction from those in the circumferential direction. In addition, it is essential for an advanced ECT probe to give a large response to a crack but low noise. To pick up a large crack signal, the pickup coils have to be located at a position where the perturbation of the secondary magnetic field due to the crack has a peak value.

For example, two geometries depicted as (b) and (c) in Fig. 37 are selected as candidates for optimal pickup coils based on the possible arrangements shown in the figure and by considering the conditions for an optimal pickup arrangements. For the plus-point coils, the differential output signal  $Z = Z_1 - Z_2$  can reduce the lift-off noise. For the four-coil pickup, by selecting the output as  $Z = (Z_1+Z_3) - (Z_2+Z_4)$  with the subscript corresponding to the coil number shown in Fig. 37, both the lift-off and inclination noise can be reduced significantly without reducing the crack signal when the ring currents are in the pattern (a) or (c) in Fig. 38 and the crack width is small. In this case, the pickup coils in the axial direction have no crack signal output due to the crack being in the circumferential direction and vice versa. Therefore, the differentiating operation does not decrease the crack signal.



**Fig. 38** Distribution patterns of the ring currents and the suitable pickup coil positions

Finally, as the three distribution patterns of the ring current as shown in Fig. 38 can be extracted for an axial crack, the typical positions to place the pickup coil are illustrated in the same figure. For a circumferential crack, similar distribution patterns and pickup positions appear but in the other direction.

Based on the above considerations, the models for the qualitative analysis of the perturbation of reaction magnetic field due to a crack will be adopted to determine the optimal arrangement of the excitation and pickup coils.

To find the optimal probe structures among the selected excitation ways and the possible pickup configuration, the most direct approach is to evaluate the detectability of each combination of these excitation and pickup coils and then to rank the possible configurations according to their detectability. To do this, the simplified evaluation method is applicable to the calculation of detectability, i.e., the S/N ratio against the lift-off noise.

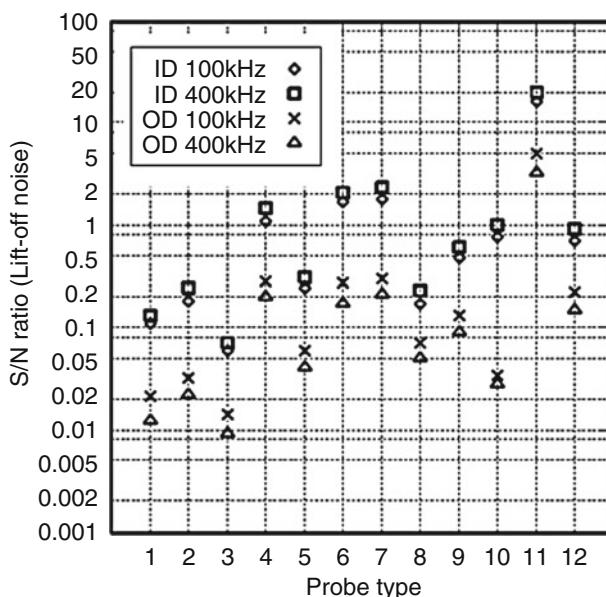
As a summary, the procedure for optimal ECT probe design can be described as follows:

1. Choose excitation structures which can induce large eddy currents in both the axial and circumferential directions.
  2. Predict the eddy current distribution from the source magnetic field using the empirical relation in Eq. (99) or Eq. (102).
  3. Determine the distribution patterns of the ring currents making use of Eq. (106) and the eddy current distribution obtained in the previous step.
  4. Decide the possible pickup locations based on the peak positions of the reaction magnetic field distribution produced by the ring currents.
  5. Select possible optimal pickup coil configurations on the basis of the ring current patterns and the following conditions:
    - (a) A position with peak magnetic field perturbation.
    - (b) Symmetry arrangement and differential output to eliminate wobbling noise.
    - (c) No significant weakening to the crack signal due to the difference operation.

6. Seek the optimal probe structure by ranking the possible combinations of excitation methods and the optimal pickups according to the S/N ratios evaluated using the simplified evaluation scheme.

This approach is feasible in the optimal design of the basic structure of ECT probes. As the quantitative accuracy of the simplified method for probe evaluation is limited, we still need to employ a high-accuracy numerical method such as the FEM-BEM code for the optimization of the detailed probe sizes.

As an example to show the efficiency of the approach, an Inconel 600 tube 1.27 mm thick and outer radius 11.115 mm is considered as the inspection object. The cracks are chosen as 5 mm in length, 0.2 mm in width, 20% of the tube wall thickness in depth and located on the inner or outer side of the tube center. The excitation frequencies are chosen as either 100 kHz or 400 kHz, while the lift-off is set as 1 mm with 0.115 mm variation. The probes are located over the center of the crack. Figure 39 shows the results of S/N ratios against lift-off noise for 12 types of probe configuration. The numbers on the horizontal scale of the figure represent the probe types which are explained in the figure, where “coil A-coil B” means a probe with excitation coil A and pickup coil B. Since the crack signals of the pancake-plus-point and four-coil-plus-point probe are zero, the results of these two probe

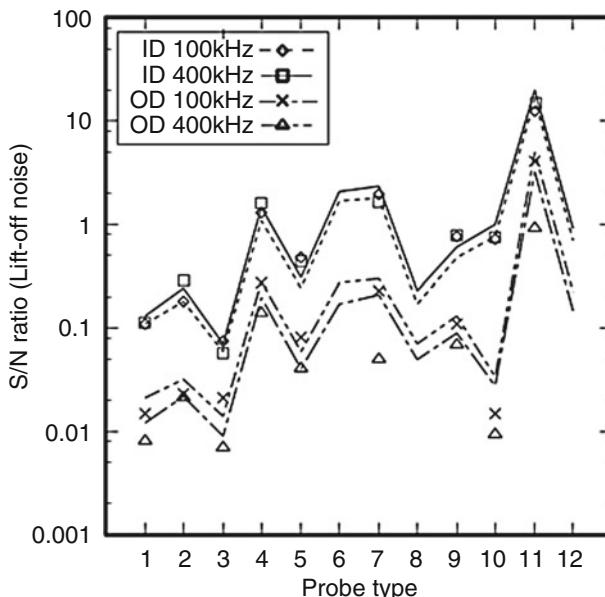


1. Self-induction pancake, 2. Mutual induction pancake, 3. New bobbin,
4. Plus-point, 5. Pancake-fourcoil, 6. Hybrid coil-fourcoil,
7. pluspoint-fourcoil, 8. Hybrid coil-pluspoint, 9. Fourcoil-fourcoil,
10. Fourcoil-fourcoil(opposite phase), 11. spiral-fourcoil, 12.spiral-pluspoint

**Fig. 39** S/N ratios of the selected probes evaluated by the simplified method (20% crack)

structures were not included in the figure. The results of an absolute pancake coil probe, a mutual induction pancake coil probe, and a new bobbin probe (excited by two bobbin-type coils, picked up by a pancake coil) are shown for comparison. From the results, it is not difficult to see that probe No. 7 (plus-point excitation-four-coil pickup) and probe No. 11 (spiral coil excitation-four-coil pickup) have better S/N ratios. In fact, the lift-off noise of probe 11 is very small when the spiral excitation coil is very long in the axial direction comparing with the diameter of the tube. The four-coil-type pickup also makes its inclination noise smaller. Based on the results introduced above, these two types of probe configurations are proposed as candidates of new high-performance ECT probes.

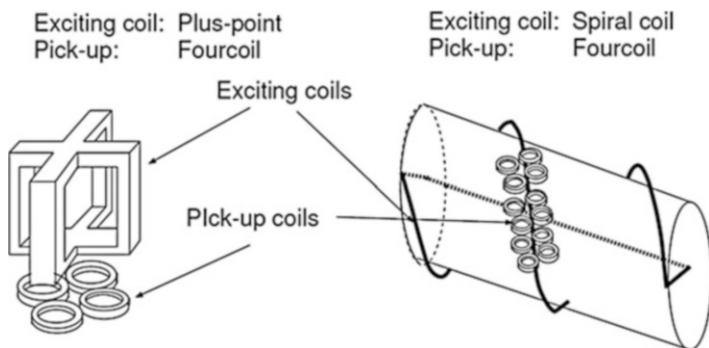
In order to validate the results of the optimization method, the hybrid code based on the FEM-BEM method was employed to perform high-accuracy detectability analyses. The comparison of the S/N ratio results is shown in Fig. 40 where the lines are the results of the simplified method and the points the S/N ratios by the FEM-BEM code. A good qualitative agreement is obtained. This result indicates that the simplified detectability analysis method provides a feasible tool for the optimal design of a new ECT probe and the two probe structures proposed above



1. Self-induction pancake, 2. Mutual induction pancake, 3. New bobbin,
4. Plus-point, 5. Pancake-fourcoil, 6. Hybrid coil-fourcoil,
7. pluspoint-fourcoil, 8. Hybrid coil-pluspoint, 9. Fourcoil-fourcoil,
10. Fourcoil-fourcoil(opposite phase), 11. spiral-fourcoil, 12.spiral-pluspoint

(lines: present method; dots: FEM-BEM method)

**Fig. 40** S/N ratios obtained by the FEM-BEM code and the simplified approach (20% crack)



**Fig. 41** Two proposed structures of optimal ECT probe

do have higher S/N ratios than the conventional high-performance ECT probes, such as the plus-point probe.

Based on the results above, two probe structures chosen as the candidates of the new high-performance probe for practical inspection are shown in Fig. 41. One of them is the probe excited by the spiral coil and measured with a four-coil pickup unit. The other one consists of a plus-point excitation coil unit and a four-coil-type pickup unit. The latter was also proposed based on another design method supported by the FEM-BEM hybrid code (Popa et al. 1997). For the actual application of these basic structures, detailed optimizations of probe sizes and applications of new techniques such as the high-speed switching are necessary to fulfill the detailed requirements of tube inspections.

## Applications of Advanced Numerical Analysis for ECT

In the previous section, the selection of probe configurations and their optimization were described in terms of forward analyses in order to understand the feature of eddy current perturbation and the effect of probe parameters on detectability. In practice, numerical methods of higher efficiency and accuracy are necessary for further development of the ECT technique, especially for defect sizing. In this section, some progress in the forward and inverse numerical schemes for ECT flaw detection and evaluation are introduced.

There are two parts in this section. The first part introduces an efficient forward analysis scheme for fast ECT signal simulation with a database strategy using pre-calculated unflawed potential field data. The second part describes the crack reconstruction method from the detected ECT signals. The reconstruction schemes based on the conjugate gradient (CG) method, a deterministic optimization category, the neural network method, and the metaheuristic optimization methods, a stochastic method, are introduced. Reconstructions of both artificial and natural cracks are performed by using measured signals to show the efficiency of the inversion

schemes. It was clarified through applications that a deterministic optimization method is more efficient for reconstruction of simple cracks, while the stochastic method is preferred for defects of complicated geometries such as a stress corrosion crack and multiple cracks.

## Forward Analysis Schemes for Fast ECT Signal Simulation

A typical numerical ECT model problem is shown in Fig. 42, where the symbol  $\Omega$  corresponds to the whole unflawed conducting object,  $\Omega_c$  is the region of crack, and  $\Omega_0$  is a selected region of regular shape that contains the crack. If we consider a slit-like crack, one can choose  $\Omega_0$  as a regular planar shape with a thickness equal to the crack width, and subdivide it into a grid of small cells, as shown in the Fig. 42. In this case, the eddy current perturbation due to a planar crack of arbitrary shape can be evaluated rapidly by using the following scheme (Chen et al. 1999).

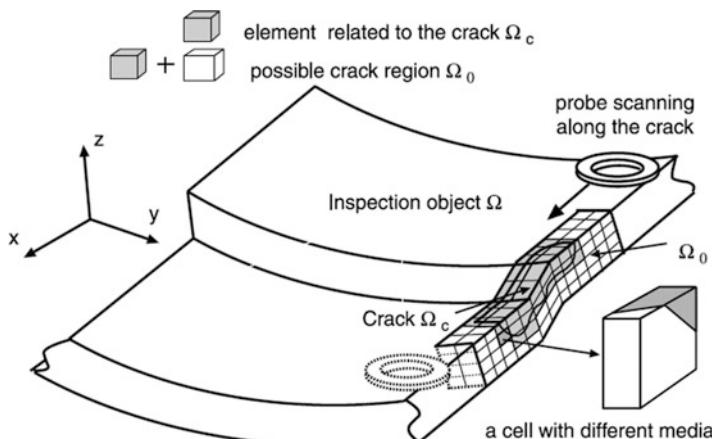
### Basic Formulation of Fast ECT Simulation Scheme

By subtracting the  $\mathbf{A} - \phi$  formulae for  $\mathbf{A}$  and  $\phi$  of a conductor with flaw present, with those for  $\mathbf{A}''$  and  $\phi''$  of the unflawed conductor, one can obtain the following governing equations of the field perturbation  $\mathbf{A}'$ ,  $\phi'$  in the conductor region

$$\frac{1}{\mu_0} \nabla^2 \mathbf{A}' - \sigma_0 (\dot{\mathbf{A}}' + \nabla \phi') = -[\sigma_0 - \sigma(\mathbf{r})] (\dot{\mathbf{A}} + \nabla \phi), \quad (114)$$

$$\nabla \cdot \sigma_0 (\dot{\mathbf{A}}' + \nabla \phi') = \nabla \cdot [\sigma_0 - \sigma(\mathbf{r})] (\dot{\mathbf{A}} + \nabla \phi), \quad (115)$$

and in air region



**Fig. 42** Numerical model of the fast scheme

$$\frac{1}{\mu_0} \nabla^2 \mathbf{A}^f = 0. \quad (116)$$

where  $\mathbf{A}^f = \mathbf{A} - \mathbf{A}^u$  and  $\phi^f = \phi - \phi^u$  are the potential perturbations due to the presence of a crack,  $\sigma_0$  is the conductivity of the host material, and  $\sigma(\mathbf{r})$  is the distribution of conductivity that is equal to 0 in the crack region and  $\sigma_0$  in the conductor. As  $[\sigma_0 - \sigma(\mathbf{r})]$  vanishes in the conducting region, after FEM-BEM discretization, Eqs. (114), (115), and (116) can be reduced to

$$\begin{bmatrix} \tilde{K}_{11} & \tilde{K}_{12} \\ \tilde{K}_{21} & \tilde{K}_{22} \end{bmatrix} \begin{Bmatrix} q_1^f \\ q_2^f \end{Bmatrix} = \begin{bmatrix} \tilde{K}_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{Bmatrix} q_1^f + q_1^u \\ q_2^f + q_2^u \end{Bmatrix} \quad (117)$$

where  $\{q^f\} = \{\mathbf{A}^f, \phi^f\}^T$ ,  $\{q^u\} = \{\mathbf{A}^u, \phi^u\}^T$  are the potentials at every node,  $[\tilde{K}]$  is the coefficient matrix of the unflawed conductor, and  $[\tilde{K}]$  is the coefficient matrix corresponding to the terms with  $[\sigma_0 - \sigma(\mathbf{r})]$ . The unknown potential vector  $\{q\}$  was divided into two parts:  $\{q_1\}$  are the potential values at the node of crack element and  $\{q_2\}$  are the remaining unknowns. As  $[\sigma_0 - \sigma(\mathbf{r})]$  vanishes in region  $\Omega - \Omega_C$ ,  $[\tilde{K}]$  is written as in Eq. (117) where the nonzero submatrix  $[\tilde{K}_{11}]$  corresponds to the elements at the crack region. Multiplying Eq. (4) left by  $[H] = [\tilde{K}]^{-1}$ , we get

$$\begin{Bmatrix} q_1^f \\ q_2^f \end{Bmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} \tilde{K}_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{Bmatrix} q_1^f + q_1^u \\ q_2^f + q_2^u \end{Bmatrix}. \quad (118)$$

The equations related to the unknowns  $\{q_1^f\}$  can be separated from Eq. (118). If we denote  $[H_{11}][\tilde{K}_{11}]$  as  $[G]$  and the unit matrix as  $[I]$ , the relation connecting the unflawed field to the field perturbation in region  $\Omega_C$  is finally deduced as follows:

$$[I - G] \{q_1^f\} = [G] \{q_1^u\}. \quad (119)$$

As the unknowns of Eq. (119) are limited to the crack region  $\Omega_C$ , calculation of the disturbed  $\{q_1^f\}$  using this equation is much faster than the conventional FEM-BEM code. This increase in speed comes from the significantly reduced number of unknowns. The information about the conductor and exciting coils are contained in the coefficient matrix  $[H_{11}]$  and vector  $\{q_1^f\}$  that can be calculated a priori as they are independent of the crack. Thus, once these fields were calculated and stored in databases, there is no need to compute them again in the actual calculation of crack signals during crack reconstruction. Consequently, the computational effort can be reduced significantly.

### Establishment of Databases of Unflawed Field Information

From the detected ECT signals, some features of a crack such as the position, rough length, and inner/outer property usually can be estimated with a classification

method or from experience. Thus, it is not difficult to choose an area containing the possible crack (such as region  $\Omega_0$  in Fig. 42) and to subdivide it into a grid of small cells. The unflawed field at  $\Omega_0$  can be calculated with the conventional FEM-BEM code by the excitation current of the probe or a unit potential located at one of the nodes in  $\Omega_0$ , respectively. Storing this unflawed information in databases, the matrices  $[H_{11}]$  and  $\{q^u\}$  of Eq. (119), which are necessary for the fast evaluation of ECT signal, can be extracted directly from the database for any crack in region  $\Omega_0$ .

### Formula for Pickup Signal

As the disturbed field  $\{q_1^f\}$  can be considered as the unflawed field induced by current dipoles in the crack region, it is convenient to calculate the pickup signals using the reciprocity theorem [16]. For a self-induction pancake coil, the impedance change is written as

$$\Delta Z = \frac{1}{I^2} \int_{\text{coil}} \mathbf{E}^f \cdot \mathbf{J}_0 dv = \frac{1}{I^2} \int_{\text{flaw}} \mathbf{E}^u \cdot (\mathbf{E}^f + \mathbf{E}^u)(\sigma_0 + \sigma(\mathbf{r})) dv, \quad (120)$$

where the term  $(\mathbf{E}^f + \mathbf{E}^u)(\sigma_0 + \sigma(\mathbf{r}))$  is the current dipole in the flaw region and  $I$  is the current per turn in the coil. The calculation of the impedance with Eq. (120) requires very little computational work because it can be assisted by the known elements of the coefficient matrices directly. The electromotive force (EMF) of a mutual inductance probe can also be evaluated by using a similar formula. In this case, the fields induced by a virtual current in the pickup coil have to be incorporated.

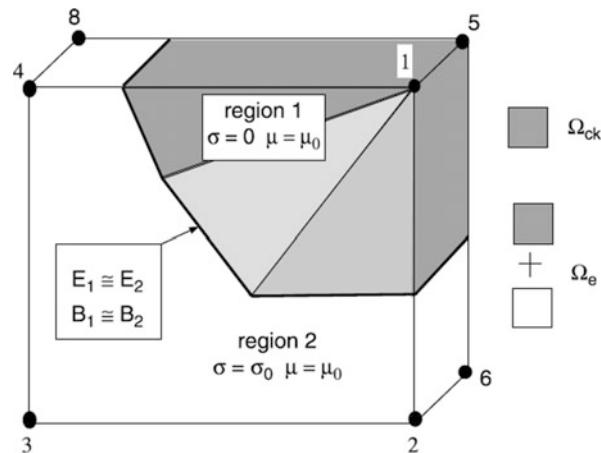
### A Special Element for FEM Discretization of Complicated Cracks

As the database has to be established using the grid of given a priori cells, only crack areas using these cells can be treated by using these databases if homogeneous conductivity is required in each FEM element. The permeability and permittivity of the nonmagnetic material are similar to free space. Therefore, the electric charge at a conductor surface is neglected for low-frequency problems. The electromagnetic fields are continuous, and the corresponding potentials  $A$  and  $\phi$  are approximately first order differentiable at the crack edge. In addition, with the fact that small grid cells are usually selected for subdividing the region  $\Omega_0$ , it is reasonable to apply the shape function of a normal FEM element to the interpolation of an unknown field even for a cell containing crack edge, i.e., with different media (see Fig. 43). Introducing such a new element will enable us to simply treat cracks of arbitrary shape in signal simulation with regular mesh. In such a case, the different material is taken into account in the element coefficient matrices, as the integration over the whole element is reduced to the part related to the crack region.

### Numerical Example of Fast-Forward Simulation for ECT Signals

In numerical calculations, the ECT signal scanned axially along the crack was evaluated for a straight tube model as an example. The size of the specimen and the ECT probe were chosen the same as in step 3 of the JSAEM benchmark models (Takagi et al. 1994). The

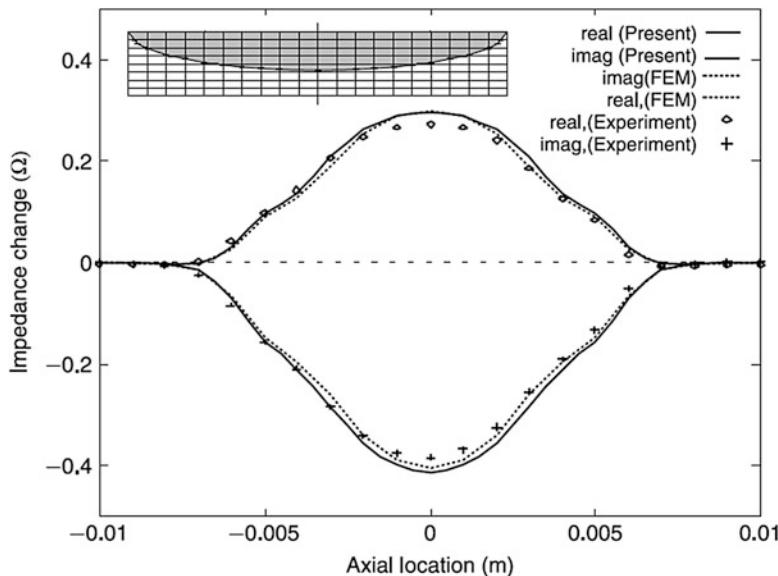
**Fig. 43** Concept of new element with different material



crack was chosen as an axial EDM notch 0.2 mm wide, though this method is suitable for more complex-shaped cracks as well. A region of dimensions  $12 \times 1.27 \times 0.2$  mm<sup>3</sup> along the axial direction was chosen as the possible crack region  $\Omega_0$  and was subdivided into  $24 \times 10 \times 1$  cells. Potentials in a region two times as long as the region  $\Omega_0$  were calculated and stored in the database for the inverse matrix, while a region with 32 mm length (three times of the length of suspicious region) was used for the database of unflawed field for the axial scan from 210 to 10 mm. The impedance results for an elliptic shaped crack are shown in Fig. 44, with the exciting frequency set at 300 kHz. The crack was located on the inner surface of the tube, with a maximum depth of 60% of the tube wall thickness and a length of 10 mm. In order to compare with the measured data, the JSAEM step 6 benchmark models (Takagi et al. 1994) were considered. As depicted in the figure, good agreement was obtained. The small difference between the two numerical results seems also to be caused by the mesh differences. The mesh used for FEM-BEM code was not as fine as that used for the present fast method.

## Techniques for Crack Reconstruction from ECT Signals

ECT inversion is a typical optimization problem, and is usually ill-posed. To solve this problem, optimization methods in both deterministic and stochastic categories are applicable. A thin crack model and the CG method have been developed based on the Green function implementation (Norton and Bowler 1993). EDM cracks in a plate were successfully reconstructed by this method. On the other hand, the trust region method for crack reconstructions was developed and applied with satisfactory results (Kojima 1997). As the conventional FEM-BEM solver was needed for the forward analysis, large simulation time was needed. Based on some other forward analysis methods and idealized crack models, reconstruction of EDM crack was also performed by many other researchers (Badics et al. 1998; Cheng et al. 1999; Tamburri and Rubinacci 2002). Work is still necessary for the practical application



**Fig. 44** Comparison of the impedance results of the present method, the FEM-BEM code, and the experiment for step 6 of benchmark problems

of these methods, especially for natural cracks and multiple cracks in different orientations. In this section, inverse algorithms developed by the authors and typical reconstruction examples are explained (Chen et al. 2004).

### Reconstruction of a Planar Crack

By using the fast-forward solver, the iteration algorithm of deterministic optimization becomes suitable for solving crack profile parameters from the measured ECT data. The reconstruction scheme for a planar crack of zero conductivity and constant opening is developed. Under these conditions, the crack reconstruction is equivalent to determine the crack edge curve. For more complicated cracks, some inversion schemes are given in the next subsection (Chen and Miya 1998a and b; Chen et al. 2000, 2006).

#### A. Basic Formulation

For prediction of the ECT signal due to a crack of realistic profile, the simulation signals must have a minimum residue compared with the measured signals. Therefore, reconstruction of the crack shape is equivalent to finding a function of the crack edge curve, which minimizes the mean square error (MSE). This functional problem can be simplified to a problem with the objective function in the Riemann space by discretizing the edge curve. To search for the optimal solution, the first-order deterministic scheme is useful in presenting the problem as the gradient of the residual error and can be simply obtained from the calculated electric field at the crack region. The major procedure of the inversion is as follows. We first define the mean-square residual as

$$\varepsilon(\mathbf{c}) = \sum_{m=1}^M |Z_m(\mathbf{c}) - Z_m^{\text{obs}}|^2, \quad (121)$$

where  $\{\mathbf{c}\}$  is the vector of the crack shape parameters.  $Z_m(\mathbf{c})$  and  $Z_m^{\text{obs}}$  are, respectively, the predicted and observed impedance signals at the  $m$ -th sampling point.  $M$  is equal to the number of scanning points.

In the present problem, as we assume that the initial values of the crack profile can be chosen as values not too far from the true ones, it is reasonable to regard the object function as the MSE defined in Eq. (121). Then, the optimal crack shape vector  $\{\mathbf{c}\}$  can be solved by minimizing the MSE iteratively

$$\{\mathbf{c}\}_n = \{\mathbf{c}\}_{n-1} + \alpha_n \{\delta\mathbf{c}\}_n, \quad (122)$$

where  $\{\mathbf{c}\}_n$  is the updated direction of the  $n$ -th iteration, which is chosen as the direction parallel to the derivative vector  $\{\partial\varepsilon/\partial C_i\}$  in the steepest descent algorithm. The  $\alpha_n$  is a step-size parameter determined as the value maximally reducing the residual. In the CG method, the vector  $\{\delta\mathbf{c}\}$  in Eq. (122) is adjusted to a new direction by considering the convergence history in order to accelerate the convergence. To calculate the gradient  $\partial\varepsilon/\partial C_i$ , it is efficient to apply the method using the adjoint field. The formulae of this method are as follows:

$$\frac{\partial\varepsilon}{\partial C_i} = 2R_e \left\{ \sum_{m=1}^M \{Z_m(\mathbf{c}) - Z_m^{\text{obs}}\} \cdot \frac{\partial Z_m(\mathbf{c})}{\partial C_i} \right\}. \quad (123)$$

The perturbation of impedance  $\delta Z$  can be expressed as follows by using the adjoint field (Norton and Bowler 1993):

$$\delta Z = \sigma_0 \int \mathbf{E} \cdot \tilde{\mathbf{E}} \delta\nu(\mathbf{r}) dV, \quad (124)$$

where  $\tilde{\mathbf{E}}$  is the adjoint electric field and  $\nu(\mathbf{r}) = (\sigma_0 - \sigma)/\sigma_0$ . For a nonconducting planar crack,  $\delta\nu(\mathbf{r})$  vanishes outside the variated crack zone due to a perturbation of the crack parameter vector  $\delta\mathbf{c}$  (along the crack edge). In this case,

$$\delta\nu(\mathbf{r}) dV = \delta\mathbf{r} \cdot \mathbf{n} ds, \quad (125)$$

where  $\mathbf{n}$  is the normal unit vector at the crack edge point  $\mathbf{r}$  and  $ds$  is the corresponding area element of the crack bottom edge. If denoting the bottom edge surface of the nonconducting planar crack as  $\mathbf{r} = \mathbf{s}(\mathbf{c}, \mathbf{t})$  ( $\mathbf{t}$  is the curve coordinate vector of the surface), one obtains

$$\delta Z_m = \sigma_0 \int_S \mathbf{E}_m \cdot \tilde{\mathbf{E}}_m \sum \frac{\partial \mathbf{s}}{\partial C_i} \cdot \mathbf{n} \delta c_i ds, \quad (126)$$

where the region  $S$  is a strip in the width of the crack opening  $h_0$  and along the crack bottom edge.

From Eq. (126) and taking into account that  $\delta Z = \sum \partial Z / \partial c_i \delta c_i$ , we obtain the following equation for the impedance derivatives:

$$\frac{\partial Z_m(\mathbf{c})}{\partial c_i} = \sigma_0 \int_S \mathbf{E}_t^m(\mathbf{r}) \cdot \tilde{\mathbf{E}}_t^m(\mathbf{r}) \frac{\partial s(\mathbf{c}, \mathbf{t})}{\partial c_i} \cdot \mathbf{n} ds. \quad (127)$$

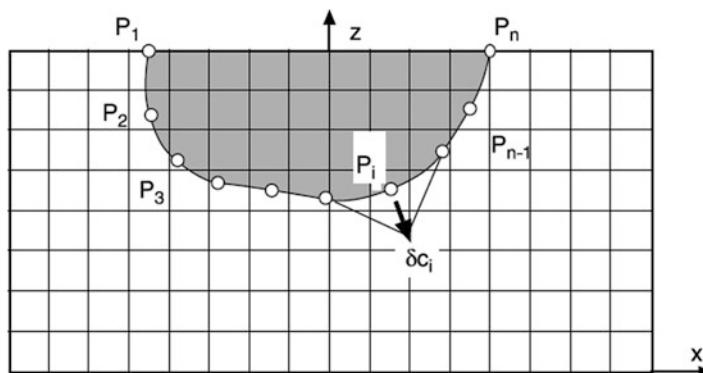
In Eq. (127), the electric field was replaced by its tangential component as no current flows across the crack bottom edge surface for a nonconducting crack. For an ECT problem of uniform material and a crack perpendicular to the object surface, the electric field is self-adjoint that simplifies Eq. (127) further. On the other hand, the step-size parameter in Eq. (122) can be calculated based on the steepest descent method by Norton and Bowler (1993):

$$\alpha_n = \frac{R_e \left\{ \sum_m (Z_m^{n-1} - Z_m^{\text{obs}}) \cdot \frac{\partial Z_m^{n-1}}{\partial \alpha_n} \right\}}{\sum_m \left| \frac{\partial Z_m^{n-1}}{\partial \alpha_n} \right|^2}. \quad (128)$$

### B. Parameterization of an EDM crack

The surface-breaking crack is of limited scope in this problem for considering further the properties of ECT. Assuming inner/outer and axial/circumferential properties are known from the phase property of ECT signal prior to the inversion, it is reasonable to express the crack with an open piecewise line as shown in Fig. 45. A constraint condition is required in this parameterization method to impose the two end points  $p_1$  and  $p_{n_c}$  located at the surface of the conductor.

In this case, the edge curve can be expressed as



**Fig. 45** The discretization of crack edge

$$x(t) = s_1(p_1, \dots, p_{n_c}, t) = x_{p_i} + \frac{x_{p_i+1} - x_{p_i}}{t_{i+1} - t_i} (t - t_i), \quad (129)$$

$$t \in [t_i, t_{i+1}] = \Gamma_i, i = 1, 2, \dots, n_c,$$

$$z(t) = s_2(p_1, \dots, p_{n_c}, t) = z_{p_i} + \frac{z_{p_i+1} - z_{p_i}}{t_{i+1} - t_i} (t - t_i), \quad (130)$$

$$t \in [t_i, t_{i+1}] = \Gamma_i, i = 1, 2, \dots, n_c - 1,$$

where  $p_1, \dots, p_{n_c}$  are  $n_c$  points equally spaced at the crack edge.  $x_{p_i}$ ,  $z_{p_i}$ , and  $t_i$  are the rectangular and curve coordinates of the point  $p_i$ , respectively. The first and last points are located at the corresponding surface point with the same  $x$  value. The variation of the edge is expressed by the variation of coordinates as follows:

$$\begin{aligned} \delta s(\mathbf{c}, \mathbf{r}) &= \delta \mathbf{r} = \delta s_1(\mathbf{c}, t) \mathbf{i} + \delta s_2(\mathbf{c}, t) \mathbf{k} \\ &= \sum_i \frac{\partial s_1(\mathbf{c}, t)}{\partial x_{p_i}} \delta x_{p_i} \mathbf{i} + \sum_i \frac{\partial s_2(\mathbf{c}, t)}{\partial z_{p_i}} \delta z_{p_i} \mathbf{k} \end{aligned} \quad (131)$$

where  $\mathbf{i}, \mathbf{k}$  denote the unit vectors along the  $x$  and  $z$  direction. Substituting Eq. (131) into Eq. (127) with  $\partial s_1(\mathbf{c}, t)/\partial x_{p_i}$  and  $\partial s_2(\mathbf{c}, t)/\partial x_{p_i}$  derived from Eqs. (130) and (131), the derivatives of the impedance  $Z_m$  are expressed in terms of the electric field through the following integrals:

$$\frac{\partial Z_m}{\partial x_{p_i}} = \sigma_0 \int_{-h_0/2}^{h_0/2} \left\{ \int_{\Gamma_{i-1}} E_{mt}^2 \frac{(t_{i+1} - t_i)(z_{p_i+1} - z_{p_i})}{(t_{i+1} - t_i)(t_{i+1} - t)} dt + \int_{\Gamma_{i-1}} E_{mt}^2 \frac{(t - t_{i-1})(z_{p_i} - z_{p_{i-1}})}{(t - t_{i-1})(t - t_{i-1})} dt \right\} dy, \quad (132)$$

$$i = 1, 2, \dots, n_c$$

$$\begin{aligned} \frac{\partial Z_m}{\partial z_{p_i}} &= \sigma_0 \int_{-h_0/2}^{h_0/2} \left\{ \int_{\Gamma_i} E_{mt}^2 \frac{(t_{i+1} - t_i)(x_{p_i+1} - x_{p_i})}{(t_{i+1} - t_i)(t_{i+1} - t)} dt + \int_{\Gamma_{i-1}} E_{mt}^2 \frac{(t - t_{i-1})(x_{p_i} - x_{p_{i-1}})}{(t - t_{i-1})(t - t_{i-1})} dt \right\} dy, \\ &\quad (133) \end{aligned}$$

$$i = 2, 3, \dots, n_c - 1$$

where the integration along the thickness direction of the crack is carried out with the variable  $y$  from  $-h_0/2$  to  $h_0/2$ . If the crack parameter  $c_i$  was chosen as the step length of the modification along the normal direction of the crack edge as shown in Fig. 45, the derivative with respect to this parameter can be obtained by imposing that point  $p_i$  move along the normal direction. Thus, the derivative with respect to  $c_i$  is equal to the partial derivative in the normal direction  $\mathbf{n}_{p_i}$ :

$$\frac{\partial Z_m}{\partial c_i} = \nabla Z_m \cdot \mathbf{n}_{p_i} = \frac{\partial Z_m}{\partial x_{p_i}} n_{x_{p_i}} + \frac{\partial Z_m}{\partial z_{p_i}} n_{z_{p_i}}. \quad (134)$$

Moreover, we only need to choose the number of points as  $n_c = 4$  to deal with a rectangular crack. In this case, the derivatives with respect to the independent crack parameters (such as the depth of the crack and the  $x$  coordinate of the two ends) can be obtained from the derivatives in Eqs. (133) and (134) using the transform matrix  $[\mathbf{T}]$  as

$$\left\{ \frac{\partial Z_m}{\partial c_i} \right\} = [\mathbf{T}] \left[ \left\{ \frac{\partial Z_m}{\partial x_{p_i}} \right\}^T, \left\{ \frac{\partial Z_m}{\partial z_{p_i}} \right\}^T \right]^T. \quad (135)$$

### C. Numerical examples of crack reconstruction from ECT signal

At first, reconstruction results for a crack of complex shape are conducted by using the impedance data computed with the FEM-BEM code as input signal. The reconstructed crack shape is shown in Fig. 46a. Crack depth, length, and shape are properly reconstructed. Figure 46b is the comparison of impedance due to crack of predicted profile and the true values. The results proved the feasibility of the proposed method for the ECT inversion of SG tube geometry.

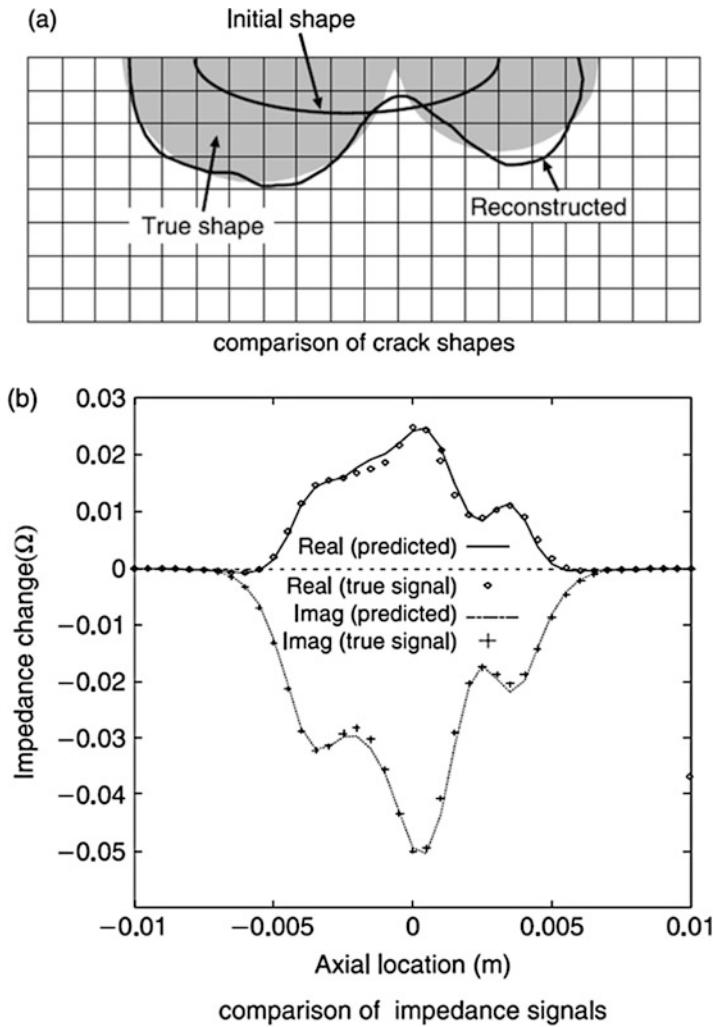
As an example of reconstruction with measured signals, the reconstruction was performed using measured signals of the JSAEM benchmark problem step 6 (Takagi et al. 1994). Figure 47a shows the inversion results for a slope-shaped crack with maximum depth of 60% of the wall thickness. Though the shape was not predicted exactly, the length and depth were obtained with good accuracy. Figure 47b shows the comparison of reconstructed and true impedance signals. Good agreement is again found from the results.

As a more difficult case, Fig. 48 shows the result for a deep crack located in a metal welding zone. A plus-point coil probe was adopted in this measurement with at a low frequency (10 kHz). In the present case, the welding bead has been removed by grinding. Though the welding part results in noise in the signal, a crack in the welding zone can be properly reconstructed even if it is as deep as 10 mm.

To consider the reconstruction of a practical crack, the effect of crack inclination and crack width on the reconstruction accuracy is investigated. In practice, an EDM crack with 20 degree inclination from the depth direction and an EDM crack of smaller width (0.05 mm) are reconstructed by using the numerical model of a 0.2 mm opening and 0 degree inclination crack. Figure 49 shows the comparison of the true and reconstructed crack profile, respectively. From these results, one can conclude that the crack profile can be properly predicted even though the crack model used for databases is somewhat different than the practical case.

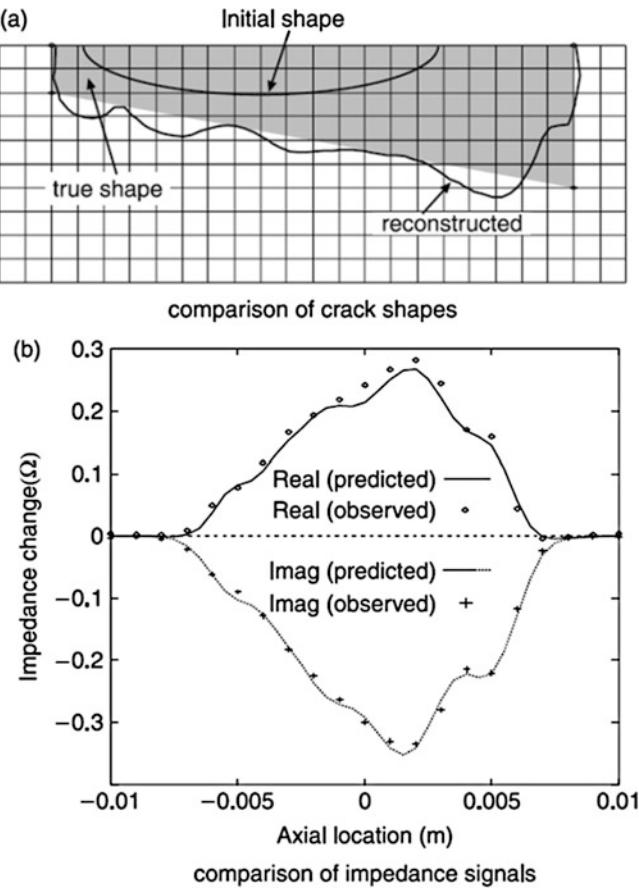
## Reconstruction of Stress Corrosion Cracks

A stress corrosion crack (SCC) has a much more complex geometry than the zero conductivity crack model. A major difference is that the crack faces may close in view of the microstructure of SCC. The closures lead to a nonvanishing conductivity inside the SCC region which complicates the numerical simulation (Chen and Miya



**Fig. 46** Reconstruction results for a crack of complex shape using the impedance calculated with the FEM-BEM code (400 kHz, tube)

1998a and b; Chen et al. 2006; Rebican et al. 2006). To reconstruct such a crack, we propose an idealized SCC model at first assuming that the closures mainly occur along the crack edge. As shown in Fig. 50, the SCC model consists of two regions separated by an inner crack edge. The conductivity of the strip along the crack edge is assumed as a value smaller than that of the base conductor, while the region in the inner part of the crack is assumed of zero conductivity. The width of the strip zone is not a constant. In this model, the effect of the crack closures is taken into account by employing both the conductivity and shape parameter of the strip region. In practice, the crack parameters are chosen as  $\sigma_1$ , the conductivity at the closure zone, and

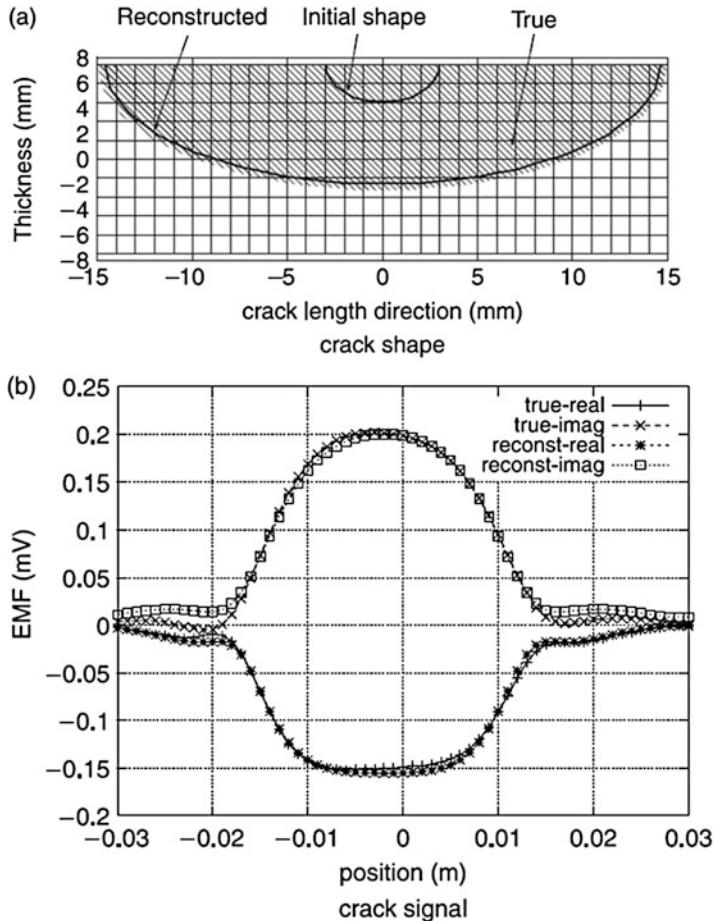


**Fig. 47** Results of reconstruction for a slope crack embedded in a conducting plate

$\mathbf{b} = \{\mathbf{b}_1, \mathbf{b}_2\}^T$ , the coordinate vector of discrete points at the two edge curves. In what follows, the inversion scheme for SCC based on this two-edge model is described.

#### A. Forward Procedures

The calculation of pickup signals for the crack with closures can be performed based on the database-type fast-forward scheme and the new element stated in the previous sections. Concerning different features of the crack parameters, however, the formulae for EDM cracks have to be modified. In the database-type fast-forward solver, matrix  $[H]$  is independent of cracks. Meanwhile, despite being affected by the crack shapes, the matrix  $[K]$  can be derived from coefficient matrices of the two cavity cracks defined by the inner and outer crack edge, respectively. The gradients with respect to the crack parameters of the idealized crack model can be obtained from the gradients corresponding to the two cavity cracks, i.e., zero conductivity. In fact, if we suppose that the shape function  $[N]$  for



**Fig. 48** Reconstruction of deep crack in a welding zone, CG method, plate is 20 mm thick

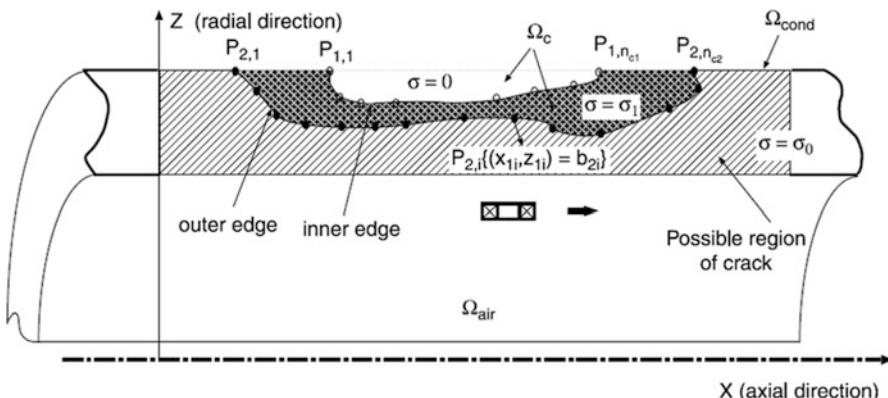
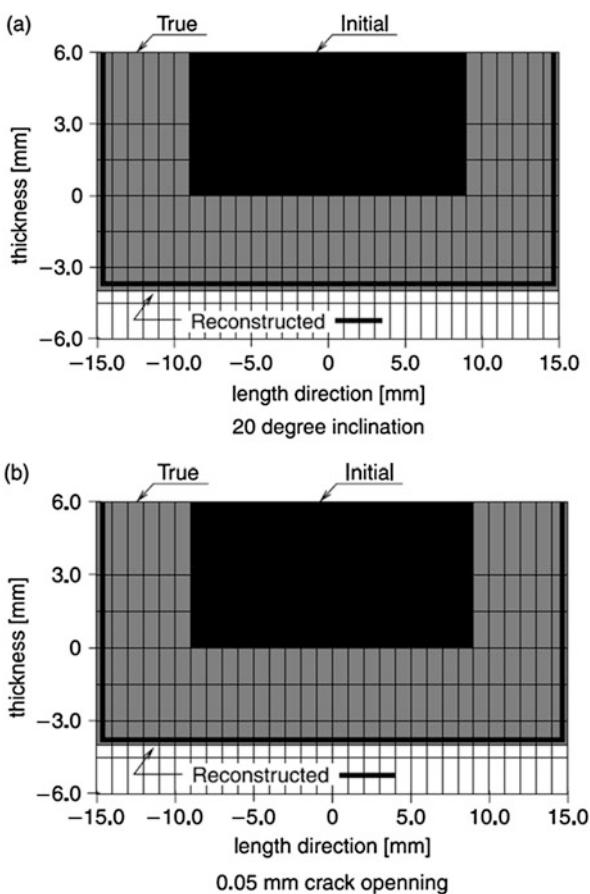
a normal finite element is still valid in the case of a distributed conductivity as applied for the new element (Chen et al. 1999), the submatrix of the element coefficients can be expressed as

$$[N_2] = \int_{\Omega_{e_1} + \Omega_{e_2} + \Omega_{e_3}} (\sigma_0 - \sigma(\mathbf{r})) [N]^T [N] dv, \quad (136)$$

$$= \int_{\Omega_{e_2} + \Omega_{e_3}} \frac{\sigma_0 - \sigma_1}{\sigma_0} \sigma_0 [N]^T [N] dv + \int_{\Omega_{e_3}} \frac{\sigma_1}{\sigma_0} \sigma_0 [N]^T [N] dv, \quad (137)$$

where  $\Omega_{e_1}$ ,  $\Omega_{e_2}$ , and  $\Omega_{e_3}$  are the regions of elements where conductivity is equal to  $\sigma_0$ ,  $\sigma_1$ , and 0, respectively. Therefore, one can write the following correlation between the practical element coefficient matrix and those for the two cavity cracks:

**Fig. 49** Effect of modeling noise (crack inclination and width difference, CG method, plate is 20 mm thick)



**Fig. 50** Numerical model of the crack with closure

$$[K] = \frac{\sigma_1}{\sigma_0} [K_1]_e + \frac{\sigma_0 - \sigma_1}{\sigma_0} [K_2]_e = \alpha_1 [K_1]_e + (1 - \alpha_1) [K_2]_e, \quad (138)$$

where  $[K_1]_e$  is an element coefficient matrix for the cavity crack surrounded by the inner edge,  $[K_2]_e$  is that defined by the outer edge, and  $\alpha_1 \equiv \sigma_1/\sigma_0$ . To predict the gradients, one can apply Eq. (127) to the two cavity cracks separately but using the electric field calculated with Eqs. (119) and (138) (true field). Denoting the gradient for the crack surrounded by the inner edge as  $\partial \bar{Z}_{1m} / \partial b_{1i}$ , that defined by the out edge  $\partial \bar{Z}_{2m} / \partial b_{2i}$  (where  $b_{1i}$  and  $b_{2i}$  are, respectively, the shape parameter for the inner and outer cavity crack;  $\bar{Z}_{1m}$  and  $\bar{Z}_{2m}$  are impedance signals calculated from the true electric field but with the formula for a cavity crack), the derivatives for the crack with closures can be expressed as

$$\frac{\partial Z_{1m}}{\partial b_{1i}} = \frac{\sigma_1}{\sigma_0} \frac{\partial \bar{Z}_{1m}}{\partial b_{1i}} = \alpha_1 [T_1] \left\{ \frac{\partial Z_m}{\partial x_{p_i}}, \frac{\partial Z_m}{\partial z_{p_i}} \right\}^T \quad (139)$$

for points located at the edge of inner side and

$$\frac{\partial Z_{2m}}{\partial b_{2i}} = (1 - \alpha_1) [T_2] \left\{ \frac{\partial Z_m}{\partial x_{p_i}}, \frac{\partial Z_m}{\partial z_{p_i}} \right\}^T, \quad (140)$$

for those at the outer edge, and  $[T_1]$  and  $[T_2]$  are the linear transform matrices. For  $\alpha_1$ , one has

$$\frac{\partial Z_m}{\partial \alpha_i} = \sigma_0 \left\{ \int_{\Omega_{ck_2}} |E_m|^2 dv - \int_{\Omega_{ck_1}} |E_m|^2 dv \right\}, \quad (141)$$

with  $\Omega_{ck_1}$  and  $\Omega_{ck_2}$  denoting the volume of the two cavity cracks, respectively. Defining  $\alpha = \sigma/\sigma_0$ , one gets

$$\frac{\partial Z_m}{\partial \alpha_1} = \frac{\partial \bar{Z}_{2m}}{\partial \alpha} - \frac{\partial \bar{Z}_{1m}}{\partial \alpha}. \quad (142)$$

## B. Inversion procedures

Compared with the case of a nonconductive crack, the most significant difference in the inverse procedure is that the crack parameters are in different types, i.e., those for the crack shape and the conductivity. The gradients with respect to these parameters usually have a much different magnitude (even in much different order of magnitude) because of the different units. Therefore, a quite different convergent speed may occur between these crack parameters if a normal algorithm of the CG method is directly applied. To tackle this difficulty, a modified scheme on the basis of the CG method is proposed.

The basic idea of this procedure is that the direction and step size at each updating step are calculated separately for parameters of different type. In practice, the iteration equation for predicting the crack parameters is taken as

$$\{\mathbf{c}_1^n, c_2^n\}^T = \left\{ \mathbf{c}_1^{n-1} + a_1^n \left\{ \frac{\partial e^{n-1}}{\partial c_1^{n-1}} \right\}, c_2^{n-1} + a_2^n \frac{\partial e^{n-1}}{\partial c_2^{n-1}} \right\}^T, \quad (143)$$

where vector  $\mathbf{c}_1 = \mathbf{b}$ , scalar  $c_2 = \alpha_1$ , and  $a_1^n$  and  $a_2^n$  are the updating step lengths at the  $n$ -th step for shape and conductivity parameters, respectively, and

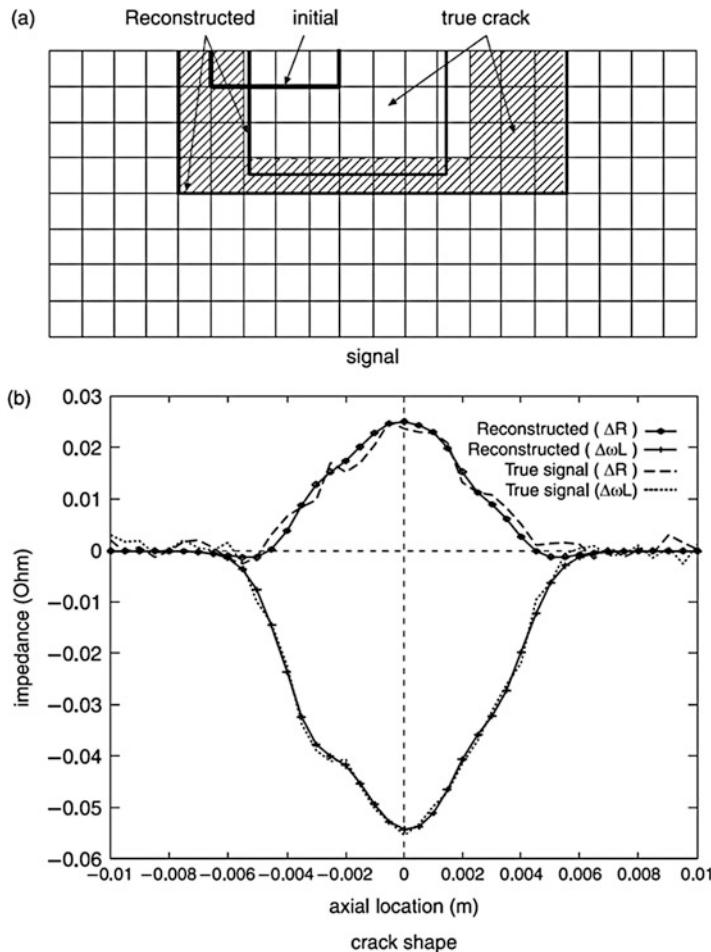
$$a_j^n = \frac{\text{Re} \left\{ \sum_{m=1}^M [Z_m^{n-1}(\mathbf{c}) - Z_m^{\text{obs}}]^* \sum_{i=1}^{N_j} \frac{\partial Z_m^{n-1}}{\partial c_{ji}} \frac{\partial e^{n-1}}{\partial c_{ji}} \right\}}{\sum_{m=1}^M \left| \sum_{i=1}^{N_j} \frac{\partial Z_m^{n-1}}{\partial c_{ji}} \frac{\partial e^{n-1}}{\partial c_{ji}} \right|^2}, j = 1, 2. \quad (144)$$

### C. Numerical examples of SCC reconstruction

Figure 51 depicts an example of impedance signals of the reconstructed crack with a comparison to the true ones (simulated data with no noise) where  $\Delta R$  and  $\Delta wL$  denote the resistance and reactance, respectively. The true shape of the crack, as shown in Fig. 51, is assumed to be an inner one with depth equal to 50% of the tube wall thickness. The reconstructed relative conductivity is 0.102 for a true value of 0.1 and an initial value of 0.5 (the conductivity of the base material was chosen as 1 MS/m). The initial values of the crack shape were chosen as the results predicted by the code for a cavity crack. Good reconstruction result is obtained for both the shape parameters and the conductivity. In addition, Fig. 52 shows a reconstruction result for a practical SCC initiated in an austenitic stainless steel pipe of 8 mm wall thickness used in a practical nuclear power plant. By using the two-edge SCC model and the CG scheme, the length and the depth of this natural SCC are satisfactorily recognized.

### Artificial Intelligent Approach for Crack Reconstruction

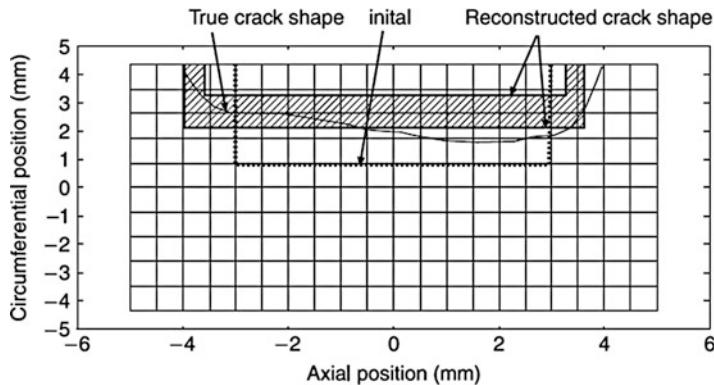
A feedforward neural network that has only one hidden layer (Yusa et al. 2000) is applied to the ECT inversion. The neural network is trained to realize the mapping between the ECT signals and the crack profiles. Considering the feature of an SCC, only surface-breaking cracks with known location and orientation are taken as reconstruction targets (Yusa et al. 2000, 2005). Based on these assumptions, a vector of integer element, which indicates the depth of the selected crack region, is proposed for the parameterization of the crack profile. In practice, a selected region of regular shape, which can be chosen directly based on the effective signal length, is subdivided into small cells and the numbers of crack cells in the depth direction are set as the values of the elements in the crack parameter vector. To generate the training data sets, the fast-forward solver was applied. For input information, the



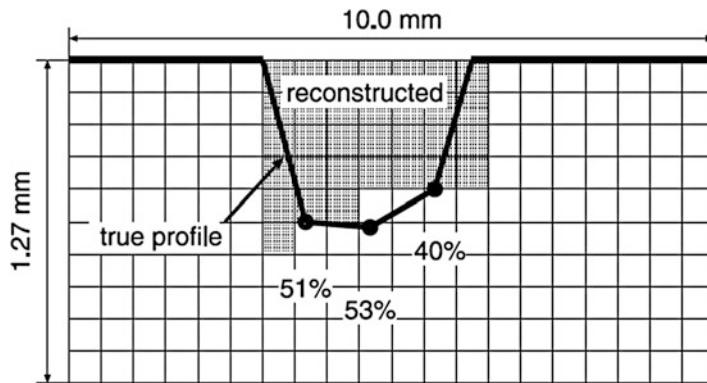
**Fig. 51** Comparison of the reconstructed and the true crack shape and signal (10% noise)

pickup signals due to cracks with their shape and conductivity distribution randomly designated are calculated for a scanning path just over the plane of the planar crack. The application of the randomly distributed conductivity to the crack cells is intended to cope with the problem of a natural crack concerning the contacts or bridges between the crack surfaces. A large number of crack signals and the corresponding crack profile vectors are calculated and stored in three databases for NN training, validating, and testing, respectively. Principal component analysis (PCA) is also applied to the raw eddy current signals as preprocessing in order to extract the features of the signal and to decrease the amount of input nodes.

Figure 53 shows the result of the neural network approach for an actual SCC detected in an SG tube (1.27 mm thickness) of a PWR nuclear power plant. The true crack shape obtained by destructive test is also illustrated. The percentage number



**Fig. 52** Reconstruction results for practical SCC, (CG method, pipe of 8 mm wall thickness)



**Fig. 53** Reconstruction results (SCC crack, neural network method, 1.27 mm tube)

shown in the figure is the reconstructed relative crack depth (ratio between reconstructed crack depth and the tube wall thickness). Good agreement is obtained again. To predict the crack shape with the trained network, almost no CPU time is consumed.

### Metaheuristic Methods: Parallel Tabu Search

Unlike the deterministic approach, metaheuristic optimization methods minimize the objective function in a stochastic way. The tabu search, genetic algorithm, and simulated annealing method belong to this category. A feature of this approach is that a global minimum can be guaranteed, although a large amount of computational resources are necessary (Yusa et al. 2003). As the simulation time required by the fast-forward solver depends on the size of the crack, starting the searching procedure of a metaheuristic method from a small initial crack and finding the optimum profile by enlarging the crack step by step are a reasonable approach. From this point of view, an algorithm based on the tabu search algorithm is more efficient in the

metaheuristic method for crack profile reconstruction from ECT signals. In addition, as parallel computation is applicable for the metaheuristic methods, parallel computation is an efficient way to reduce the simulation time further. The major procedure of the tabu search is as follows:

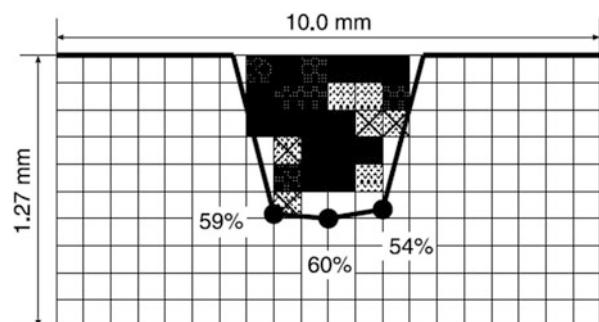
1. Randomly select one of the cells that face the surface of the target plate as an initial crack (solution candidate). Generate initial cracks in this manner.
2. Calculate residual function  $e$  of each candidate by using  $n$  CPUs in parallel.
3. If the terminate condition is satisfied, output the best solution in the history as the final prediction.
4. If not, add all of the solution candidates evaluated in step 2 to the tabu list.
5. Copy the best solution candidate in the history into all of the  $n$  solution candidates. Generate  $n$  new solution candidates for the next iteration by slightly updating the present ones. If a new solution candidate is already in the tabu list, regenerate it.
6. Return to step 2.

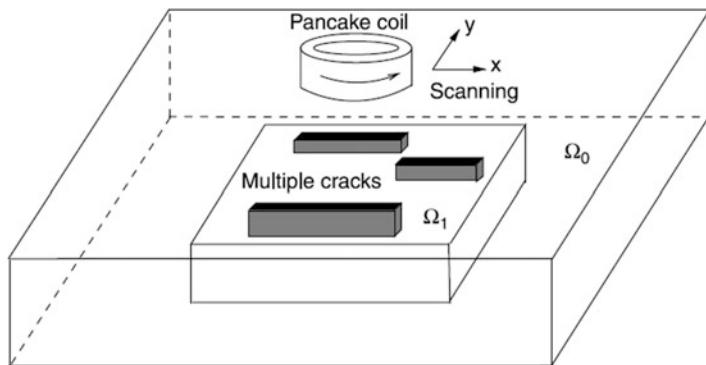
Figure 54 shows a reconstruction result of the tabu search algorithm for the same problem as used for the NN approach. The percentage number shown in the figure is the reconstructed relative crack depth. Better results are obtained. However, the simulation needs much more computational resource than the CG method and the artificial neural network.

### Reconstruction of Multiple Cracks

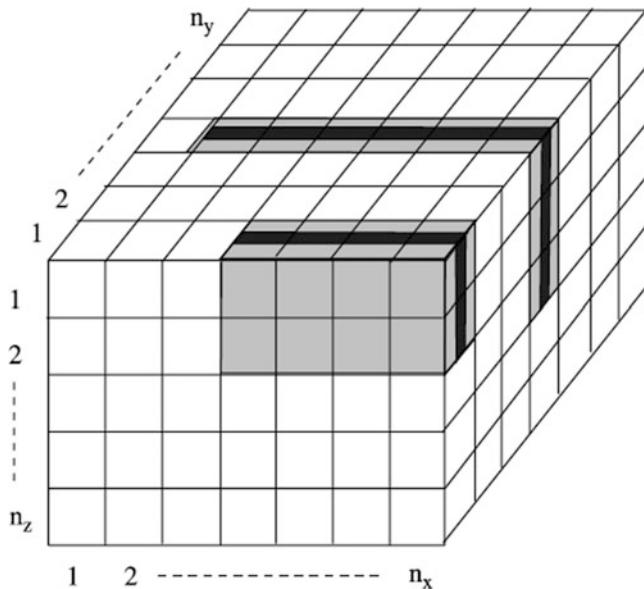
At first, we give an effective rack model for multiple slit reconstruction. All cracks are assumed to be oriented in the same direction. To consider a specimen with EDM artificial cracks, cells that form cracks have zero conductivity, and cells outside cracks have conductivity of the base material (Rebican et al. 2004, 2006). The cracks are searched in a selected region  $\Omega_1$  chosen from the whole domain based on the features of the ECT signals. The domain  $\Omega_1$  is subdivided into  $n_x \times n_y \times n_z$  cells that form geometrically the possible cracks. Figure 55 shows the configuration of the considered multiple crack problem. Since the width of an EDM notch does not significantly affect the ECT signals, it is assumed as a given value of 0.2 mm. The

**Fig. 54** Reconstruction results (SCC crack, tabu search method, 1.27 mm tube)



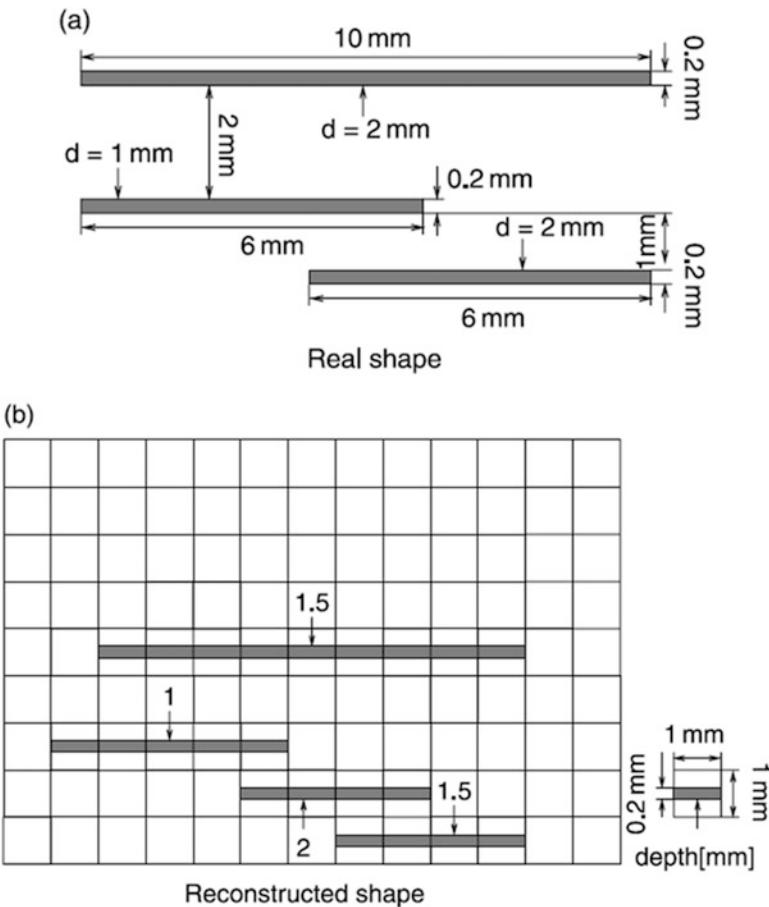


**Fig. 55** Configuration of multiple crack problem



**Fig. 56** Defect region division for multiple slits

cracks did not fill up the cells completely in the direction perpendicular to the cracks (see Fig. 56). Though the boundary profiles of defects did not coincide with the cells division, the finite element mesh needed to simulate signals was chosen based on the selected crack width. The crack parameters (vector  $c$ ) are defined as the number of starting cells of the crack, the number of final cells of the crack, and the crack depth (Fig. 56). In the following, the reconstruction scheme for the slits model, and validation using measured signals are given. The database-type forward solver was upgraded for the multiple crack case. For the inverse analysis, the tabu search technique was selected.



**Fig. 57** Reconstruction results of three parallel slits: left, true; right, reconstructed

#### A. Forward analysis

The fast-forward FEM-BEM solver using database [12] was first upgraded for the computation of the ECT signals due to multiple cracks. For the selected 3D possible defect region, the field values for cells with the same width as the cracks are required in the database. As only the fields where the sources are set in the center of the possible defect region need to be calculated, a coarse mesh was used in the regions far from the center region. As the field in the far regions does not change sharply, a linear interpolation of the field can give sufficient precision.

#### B. Reconstruction method using a tabu search

Tabu search is applied for reconstruction of multiple cracks. While the basic tabu search explores the search space of all possible solutions by a sequence of moves, tabu search here utilizes an iterative local search to generate new solution candidates because evaluation of all possible solutions in a neighborhood is too

time-consuming. This algorithm, which is based upon the iterative local search superimposed with tabu rules, is termed simply as tabu search in this work. Therefore, the difference between the tabu search and the iterative local search is in the generation of solution candidates for the next iteration step. Tabu rules adopted in this study are the two most fundamental ones: Explicit and attribute tabu strategies. The explicit tabu strategy stores defect profiles evaluated during the inversion procedure, and the generation of a evaluated defect is prohibited thereafter; the attribute tabu strategy prohibits a certain movement for the given iteration periods, if the reverse of the movement has recently improved the error. The stop condition of the iterative process is whether the error function is small enough or the maximum number of iterations is reached.

### C. Numerical examples for multiple cracks reconstruction

The raw C-scan measured data from single and multiple groups of slits were calibrated and used for the crack reconstruction. Figure 57 Shows some results for multiple parallel slits (Group B) in comparison with the real shape from measured data. From the results, the measured data are used from two groups of slits (A and B) and are modeled with approximation due to the restrictions of the proposed model (the minimum distance between slits was 0.8 mm). The tabu search is able to find with a good accuracy the main features of multiple slits, location, length, and depth, without any knowledge about the number of cracks and distance between them in the searched 3D region.

---

## References

- Achenbach JD (2000) Quantitative nondestructive evaluation. *Int J Solids Struct* 37:1–27
- Atherton DL (1995) Remote field eddy current inspection. *IEEE Trans Magn* 31(6):4142–4147
- Auld BA, Moulder JC (1999) Review of advances in quantitative eddy current nondestructive evaluation. *J Nondestruct Eval* 18:3–36
- Burke SK (1986) Impedance of a horizontal coil above a conducting half space. *J Phys D Appl Phys* 19:1159–1173
- Bowler J (1987) Eddy current calculations using half-space Green's functions. *J Appl Phys* 61 (3):833–839
- Badics Z, Pavo J, Komatsu H (1998) Fast flaw reconstruction from 3D eddy current data. *IEEE Trans Magn* 34:2823–2828
- Chari MVK (1974) Finite element solution of the Eddy current problem in magnetic structures. *IEEE Trans Power Appar Syst* 93(1):62–72
- Chen Z, Aoto K, Miya K (2000) Reconstruction of cracks with physical closure for signal of eddy current testing. *IEEE Trans Magn* 36:1018–1022
- Cochran A, Carr C (1995) Recent progress in SQUIDs as sensors for electromagnetic NDE. *Studi Appl Electromagn Mech* 8:75–86
- Cocco VS, Drunan GV, Sharp FL (1986) Eddy current manual: volume 1: test method, NUC-CAN-AECL-7523 Rev.1, Atomic Energy of Canada Limited
- Chen Z, Miya K (1998a) ECT inversion using a knowledge based forward solver. *J Nondestruct Eval* 17(3):167–175
- Chen Z, Miya K (1998b) A new approach for optimal design of eddy current probes. *J Nondestruct Eval* 17(3):105–116

- Cheng W, Miya K, Chen Z (1999) Reconstruction of cracks with multiple eddy current coils using a database approach. *J Nondestruct Eval* 18:149–160
- Chen Z, Miya K, Kurokawa M (1997) A distinctive featured optimization approach for ECT probes. *Rev Prog Quant Nondestr Eval* 16:989–996
- Chen Z, Miya K, Kurokawa M (1999) Rapid prediction of eddy current testing signals using A- $\phi$  method and database. *NDT&E Int* 32:29–36
- Chen Z, Rebican M, Miya K, Takagi T (2005) 3D simulation of remote field ECT by using Ar method and a new formula for signal calculation. *Res Nondestr Test* 16:35–53
- Chen Z, Rebican M, Yusa N, Miya K (2006) Fast simulation of ECT signal due to a conductive crack of arbitrary width. *IEEE Trans Magn* 42:683–686
- Chen Z, Takashima H, Miya K (2004) A hybrid database approach for simulation of remote field eddy current testing signals. *Int J Appl Electromagn Mech* 19:219–223
- Chen Z, Xie S, Li Y (eds) (2015) Electromagnetic nondestructive evaluation (XVIII). IOS Press, Amsterdam
- Chen Z, Xie S, Li W (2011) Reconstruction of stress corrosion crack with multi-frequency ECT signals. Paper presented at the 8th international conference on flow dynamics, Tohoku University, Sendai, 8–12 Oct 2011
- Chen H, Xie S, Zhou H, Chen Z (2014) Numerical simulation of magnetic incremental permeability for ferromagnetic material. *Int J Appl Electromagn Mech* 45:379–386
- Chen Z, Yusa N, Miya K (2008) Enhancements of ECT techniques for quantitative nondestructive testing of key structural components of nuclear power plants. *Nucl Eng Des* 238(7):1651–1656
- Chen Z, Yusa N, Miya K (2009) Some advances in numerical analysis techniques for quantitative electromagnetic nondestructive evaluation. *Nondestr Test Eval* 24(1):69–102
- Chen Z, Yusa N, Miya K (2004b) Advanced MFLT for detecting far side defects in a welding part of an austenitic stainless steel plate. *Int J Appl Electromagn Mech* 19:527–532
- Chen Z, Yusa N, Miya K (2004c) Inversion techniques for eddy current NDE using optimization strategies and a rapid 3D forward simulator. *Int J Appl Electromagn Mech* 20:179–187
- Chen Z, Yusa N, Miya K (2006b) Reconstruction of natural stress corrosion crack in coolant tubes from eddy current testing signals. *Stud Appl Electromagn Mech* 26:197–204
- Davis J (1996) Nondestructive evaluation and quality control. ASM Int, Materials Park
- Dodd CV (1977) The use of computer modelling in Eddy current testing. *Res Tech Nondestr Test* 2:429–479
- Dodd C, Deeds W (1968) Analytical solutions to eddy-current probe-coil problems. *J Appl Phys* 39:2829–2838
- Demerdash NA, Nehl TW (1978) An evaluation of the methods of finite elements and finite differences in the solution of nonlinear electromagnetic fields in electrical machines. *IEEE Trans Power Appar Syst* 98(1):74–87
- Forster F (1959) Nondestructive testing handbook, vol 2, 1st edn. Am Soc Nondestr Test, Columbus, pp 36–42
- Fuller E (2006) Steam generator integrity assessment guidelines, Rev 2, Report No.101298. Electric Power Research Institute, Palo Alto
- Fukutomi H, Takagi T, Tani J (1998) Three-dimensional finite element computation of a remote field eddy current technique to non-magnetic tubes. *J JSAEM* 6:343–349
- Hagemaijer DK (1985) Eddy-current standard depth of penetration. *Mater Eval* 10(43):1438–1454
- Harvey ED (1995) Eddy current testing: theory and practice, ASNT reference manual. The American Society for Nondestructive Testing, Columbus
- Hayt HW (2006) Engineering electromagnetics. McGraw-Hill, New York
- Haus H, Melcher J (1989) Electromagnetic fields and energy. Prentice-Hall, Englewood Cliffs
- Hellier C (2003) Handbook of nondestructive evaluation. McGraw-Hill, New York
- Hughes DV (1879) Induction-balance and experimental researches therewith. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 8:50–56
- Huang L, He R, Zeng Z et al (2012) An extended iterative finite element model for simulating eddy current testing of aircraft skin structure. *IEEE Trans Magn* 48(7):2161–2165

- He Y, Luo F, Pan M (2010) Pulsed eddy current technique for defect detection in aircraft riveted structures. *NDT & E Int* 43(2):176–181
- Hernandez JH, Pacheco ER, Caley F (2012) Rapid estimation of artificial near-side crack dimensions in aluminium using a GMR-based eddy current sensor. *NDT&E Int* 51(1):94–100
- He DF, Shiwa M, Jia JP (2011) Multi-frequency ECT with AMR sensor. *NDT & E Int* 44 (5):438–441
- Huang H, Sakurai N, Takagi T (2003) Design of an eddy-current array probe for crack sizing in steam generator tubes. *NDT&E Int* 36:515–522
- Higashi M, Tokuhisa K, Kurokawa M (2008) Development of Eddy current testing technique for PWR Vessel's dissimilar metal weld. *J JSNDI* 57(5):232–235
- Hashimoto M, Uesaka M, Miya K (1993) Development of magnetic field visualization system using hall device array probe. *Sens Mater* 4(6):313–321
- IAEA Training Course Series (2011) Eddy current testing at level 2: manual for syllabi, IAEA TEC-DOC-628 Rev2. International Atomic Energy Agency, Vienna
- Ishibashi K (1995) Eddy current analysis by the boundary integral method. *IEEE Trans Magn* 31:1500–1503
- JSAEM report (1997) Report on Advanced ECT technique, JSAEM-R-9601
- Janousek L, Capova K, Gombarska D et al (2009) Recent Trends and Developments in Eddy Current Non-Destructive Sensing. Czech Republic, Cheb
- Jomdecha C, Cai W, Xie S, Chen Z (2018) Analysis of magnetic flux perturbation due to conductivity variation in equivalent stress-corrosion crack. *Int J Appl Electromagn Mech* 2018, 59. <https://doi.org/10.3233/JAE-171140>
- Janousek L, Chen Z, Yusa N (2005) Excitation with phase shifted fields-enhancing evaluation of deep cracks in eddy-current testing. *NDT&E Int* 38:508–515
- Jogschies L, Klaas D, Kruppe R et al (2015) Recent developments of Magnetoresistive sensors for industrial applications. *Sensors* 15:28665–28689
- Jander A, Smith C, Shneider R et al (2005) Magnetoresistive sensors for nondestructive evaluation. Paper presented at the 12th International Symposium of Nondestructive Evaluation for Health Monitoring and Diagnostics, San Diego, 8–12 Mar 2005
- Kojima F (1997) Numerical scheme for reconstruction of crack shape in SG tubing by using FEMBEM hybrid code and inverse analysis. *Trans JSME* 63:2650–2656
- Kurokawa M (1997) Development of new eddy-current testing probe. *Stud Appl Electromagn Mech* 12:177–183
- Kreutzbruck MV, Krause HJ (2002) HTs squids for the nondestructive evaluation of composite structures. *Physica C: Superconduct* 368(1–4):70–79
- Kosmas K, Sargentis CH, Tsamakis D (2005) Non-destructive evaluation of magnetic metallic materials using hall sensors. *Sens Actuators A* 161(1–2):359–362
- Kim J, Yang G, Udma L (2010) Classification of pulsed eddy current GMR data on aircraft structures. *NDT&E Int* 43:141–144
- Li Y, Bei Y, Li D, Chen Z (2016) Gradient-field pulsed Eddy current probes for imaging of hidden corrosion in conductive structures. *Sens Actuators A* 238:251–265
- Ludwig R, Dai X (1990) Numerical and analytical modeling of pulsed eddy currents in a conducting half-space. *IEEE Trans Magn* 26:299–307
- Libby HL, Wandling CR (1970) Eddy current multi-parameter test for tube flaws in support region. BNWL-1468
- Li W, Xie S, Chen Z (2013) Reconstruction of stress corrosion cracks using signals of pulsed eddy current testing. *NDT&E Int* 28(2):145–154
- MacMaster RC (1963) Nondestructive testing handbook. The Ronald Press, New York
- Miya K (1995) Analytical electromagnetics and electromagnetic structures. Yokendo Press, Tokyo
- Mottl Z (1990) The quantitative relations between true and standard depth of penetration for air-cored probe coils in eddy current testing. *NDT&E Int* 23:11–18
- Martin JG, Gil JG, Sanchez EV (2011) Non-destructive techniques based on Eddy current testing. *Sensors* 11(3):2525–2565

- Mook G, Hesse O, Uchanin V (2007) Deep penetrating Eddy currents and probes. Mater Test 49:258–264
- Maeda K, Shimone J, Harada Y (1997) Optimization of transmit-receive coils for ECT probe with use of the 3-D FEM code. Electromagnetic nondestructive evaluation. IOS Press, Amsterdam. pp 199–206
- National Research Council (1997) Aging of U.S. Air Force aircraft. National Academy Press, Washington, DC
- Norton SJ, Bowler JR (1993) Theory of eddy current inversion. J Appl Phys 73:501–512
- Popa RC, Miya K, Kurokawa M (1997) Optimized eddy current detection of small cracks in steam generator tubing. J Nondestruct Eval 16(3):161–173
- Recommended Practice NO. SNT-TC-1A (2016) Personnel qualification and certification in non-destructive testing. American Society for Nondestructive Testing, Columbus
- Richard W (1996) Rules for In-service Inspection of Nuclear Power Plant Components, ASME boiler and pressure vessel code section XI. The American Society of Mechanical Engineers
- Ripka P (2003) Advances in fluxgate sensors. Sensors Actuators A 106(1–3):8–14
- Rothwell E, Cloud M (2001) Electromagnetics. CRC Press, Boca Raton
- Rebican M, Chen Z, Yusa N et al (2005) Investigation of numerical precision of 3D RFECT signal simulation. IEEE Trans Magn 41:1968–1971
- Rebican M, Chen Z, Yusa N (2006) Shape reconstruction of multiple cracks from ECT signals by means of a stochastic method. IEEE Trans Magn 42:1079–1082
- Reis D, Lambert M, Lesselier D (2002) Eddy-current evaluation of three-dimensional defects in a metal plate. Inverse Probl 18:1857–1871
- Ramos H, Postolache O, Alegria F (2009) Using the skin effect to estimate cracks depths in metallic structures. IEEE Instr & Meas Tech 21(12):1361–1366
- Pelkner M, Pohl R, Erthner T (2015) Eddy Current Testing with High-Spatial Resolution Probes Using MR Arrays as Receiver. Paper presented at the 7th International Symposium on NDT in Aerospace, Bremen, 16–18 Nov 2015
- Ramos HG, Ribeiro AL (2014) Present and future impact of magnetic in NDE. Procedia Eng 86:406–419
- Postolache O, Ribeiro A, Ramos H (2012) Uniform Eddy current probe based on GMR sensor Array and image processing for NDT. IEEE Int Instr Measur Tech 8443(3):458–463
- Ripka P, Vopalensky M, Platil A (2003) AMR magnetometer. J Magn Magn Mater 254:639–641
- Rebican M, Yusa N, Chen Z (2004) Reconstruction of multiple cracks in an ECT round-robin test. Int J Appl Electromagn Mech 19:399–404
- Sabbagh H, Sabbagh L (1986) An eddy-current model for three-dimensional inversion. IEEE Trans Magn 22:282–291
- Tai C (1971) Dyadic green functions in electromagnetic theory. Oxford University Press, Oxford
- Tumanski S (2007) Induction coil sensors- a review. Meas Sci Technol 18(3):R31–R46
- Thompson DO, Chimenti DE (1992) Review of Progress in quantitative nondestructive evaluation, vol 15A. Plenum Press, New York, pp 781–788
- Takagi T, Huang H, Fukutomi H (1998) Numerical evaluation of correlation between crack size and Eddy current testing signal by a very fast simulator. IEEE Trans Magn 34(5): 2581–2584
- Takagi T, Hashimoto H, Fukutomi H (1994) Benchmark models of eddy current testing for steam generator tube: experiment and numerical analysis. Int J Appl Electromag in Materials 4 (5):149–162
- Tegopoulos JA, Kriegis EE (1985) Eddy current in linear conducting media. Elsevier, Amsterdam
- Tian GY, Li Y, Mandache C (2009) Study of lift-off invariance for pulsed Eddy-current signals. IEEE Trans Magn 45:184–191
- Tamburrino A, Rubinacci G (2002) A new non-iteration inversion method for electrical resistance tomography. Inverse Probl 18:1809–1829
- Takagi T, Uesaka M, Miya K (1997a) Electromagnetic NDE research activities in JSAEM. Stud Appl Electromagn Mech 12:9–16

- Takagi T, Uesaka M, Miya K (1997b) Electromagnetic NDE research activities in JSAEM, electromagnetic nondestructive evaluation. IOS Press, Amsterdam. pp 9–16
- Udpa S, Moore P (2004) Nondestructive testing handbook: electromagnetic testing, 3rd edn. American Society for Nondestructive Testing, Columbus, Ohio
- Xie S (2012) Quantitative Nondestructive Evaluation of Pipe Wall Thinning Using Pulsed Eddy Current Testing. Dissertation, Tohoku University
- Xie S, Chen Z, Takagi T et al (2011) Efficient numerical solver for simulation of pulsed Eddy current testing signals. *IEEE Trans Magn* 47:4582–4591
- Yusa N, Chen Z, Miya K (2000) Quantitative profile evaluation of natural crack in steam generator tube from eddy current signals. *Int J Appl Electromagn Mech* 12:139–150
- Yusa N, Chen Z, Miya K (2003) Large scale parallel computation for the reconstruction of natural stress corrosion cracks from eddy current testing signals. *NDT&E Int* 36:449–459
- Yusa N, Chen Z, Miya K (2005) Sizing of stress corrosion cracks in piping of austenitic stainless steel from eddy current NDT signals. *Nondestruct Test Eval* 20:103–114
- Yusa N, Perrin S, Mizuno K (2007a) Eddy current inspection of closed fatigue and stress corrosion cracks. *Meas Sci Technol* 18:3403–3408
- Yusa N, Perrin S, Mizuno K (2007b) Numerical modeling of general cracks from the viewpoint of eddy current simulations. *NDT&E Int* 40:577–583
- Zenglu S, Tsutomu Y, Hideki S (2011) Detection of damage and crack in railhead by using eddy current testing. *J Electromagn Anal Appl* 3:546–550



# High-Frequency Eddy Current Techniques 20

Susanne Hillmann, Martin H. Schulze, and Henning Heuer

## Contents

Introduction .....	730
Basics of High-Frequency Eddy Current Method .....	730
Instrumentation .....	732
Typical Applications .....	733
High-Frequency Eddy Current Techniques for Application to Carbon Composites .....	733
High-Frequency Eddy Current Techniques for Ceramic Composites .....	740
High-Frequency Eddy Current Techniques for Characterization of Materials with Low Electrical Conductivity .....	741
High-Frequency Eddy Current Techniques of Polymers and Dielectrics .....	752
Summary .....	754
References .....	756

## Abstract

High-Frequency Eddy Current Techniques (HFEC) refer to special Eddy Current Techniques which work in the frequency range of around 100 kHz to 100 MHz. The techniques are based on the normal Eddy Current technique. Using a coil, an electromagnetic field is induced in the surface of the component, which has to be evaluated. The electromagnetic field induces circling currents, the so called eddy currents. Depending on the electrical properties of the material in the area of this

---

S. Hillmann (✉)

Accredited Test Lab, Fraunhofer IKTS, Dresden, Germany

e-mail: [Susanne.hillmann@ikts.fraunhofer.de](mailto:Susanne.hillmann@ikts.fraunhofer.de)

M. H. Schulze

Eddy Current Methods, Fraunhofer IKTS, Dresden, Germany

e-mail: [martin.schulze@ikts.fraunhofer.de](mailto:martin.schulze@ikts.fraunhofer.de)

H. Heuer

Systems for Testing and Analysis, Fraunhofer IKTS, Dresden, Germany

e-mail: [henning.heuer@ikts.fraunhofer.de](mailto:henning.heuer@ikts.fraunhofer.de)

exciting field, a second electromagnetic field is induced, which is directed perpendicular to the exciting field. By analyzing the complex impedance of the measurement coil, variations of the electric properties of the medium can be analyzed. In the higher frequency range of HFEC, special sensors and hardware filters must be used. HFEC is better suited for applications in materials with very low electrical conductivity, such as titanium alloys, nickel-based superalloys and similar materials, Carbon composites, ceramic composites, and polymers and lossy dielectrics may be evaluated as well. With HFEC, it is possible to reach an improvement in the measurement effect as well as better lateral resolution with matching penetration depths for these applications in comparison to lower eddy current frequencies. Additionally, the measurement effect is not only influenced by inductive processes in the area affected by the eddy currents but also by capacitive effects and an increase in displacement currents, which increases the gathered information as well. Regarding the applications, measurement process, sensors and hardware, different specifics must to be considered.

---

## Introduction

This chapter gives an introduction to high-frequency eddy current technology as a special form of the eddy current method. Specific characteristics of the technique are explained and special approaches are described.

---

## Basics of High-Frequency Eddy Current Method

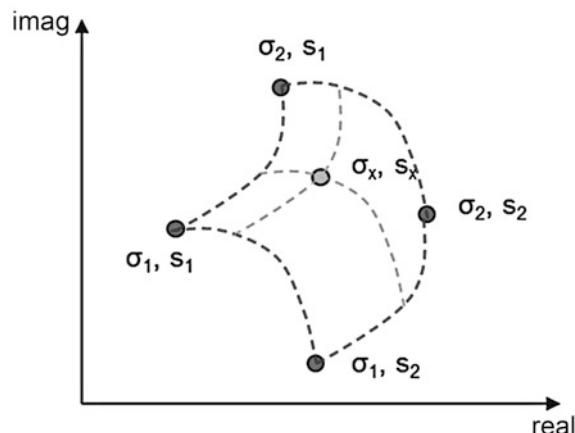
The high-frequency eddy current method (HFEC) is based on the normal eddy current method but works in the higher frequency range of around 100 kHz to 100 MHz. The basic principle is similar to classic eddy current techniques. Using a coil, an electromagnetic field is induced in the surface of the component, which needs to be analyzed. As a result circling currents are induced, the so called Eddy Currents. Depending on the electrical properties of the material in the area of this exciting field, a second electromagnetic field is induced, which is directed perpendicular to the exciting field. By analyzing the complex impedance of the measurement coil, these variations of the electric properties can be analyzed. For the higher frequency range, special electronics and sensors are needed, which is described in the following paragraph. Additionally, a more specific calibration of the system is needed due to an increasing influence of lift-off variations and increasing curved lift-off – and the conductivity locus in complex impedance plane at high frequencies. The lift-off-sensitivity is high in that frequency range that careful manual connection is not sufficient. Even in the case of very careful application, small variations in Lift-Off in the micrometer range can influence the measurements.

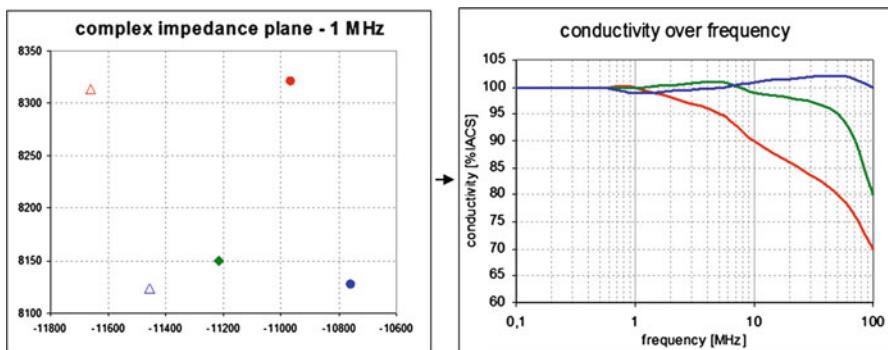
One approach is the so called four-point-calibration. Calibration blocks with known and preferably constant electrical conductivity were used, one block with a little bit lower conductivity and one block with a little bit higher conductivity for all measured samples in the sample sets. The calibration blocks have to be very homogeneous, without any layer on the surface with varying conductivity. The conductivity range of the block should frame the summarized electrical conductivity of the samples, including the influence of their layer. Additionally, it is important that the conductivity range between the calibration blocks and the sample is very small, due to the Lift-Off locus and conductivity locus in complex impedance plane are increasingly curved in the high frequencies, as Fig. 1 demonstrates. This fact influences the measurements increasingly but can be alleviated by careful calibration.

These calibration blocks were measured with and without an exactly known lift-off with a complete frequency sweep, resulting in an area in the complex impedance plane that is limited by four calibration points for each measurement frequency, see Fig. 2. The results of the measurements on the unknown samples should be located inside this area. Due to interpolation, the apparent conductivity and lift-off for each sample and each frequency can be calculated. With this procedure, frequency profiles of the so-called “apparent eddy current conductivity” (AECC) can be performed. Figure 19 demonstrates this four-point-calibration

HFEC is better suited for application on materials with very low electrical conductivity, such as titanium alloys, nickel-based superalloys and similar materials, carbon composites, ceramic composites, and polymers and dielectrics as well. It is possible to reach an increasing measurement effect as well as a better lateral resolution with matching penetration depths for these applications in comparison to lower eddy current frequencies. Additionally, the measurement effect is not only influenced by inductive processes but also influenced by capacity effects and a growth of displacement currents, which increases the gathered information as well. Regarding the application, measurement process, sensors and hardware, different specifics have to be considered (Abu-Nabah et al. 2009; Hillmann et al. 2010; Heuer et al. 2009).

**Fig. 1** Lift-off locus and conductivity locus curves in the complex impedance plane





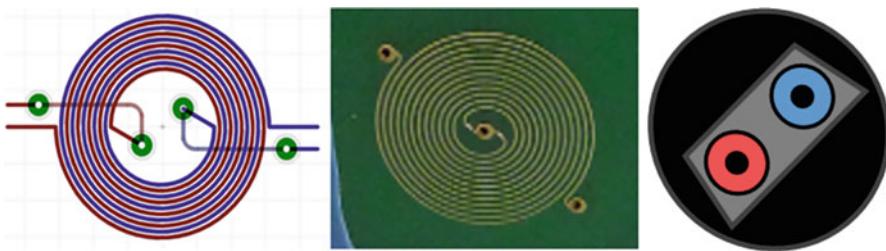
**Fig. 2** Illustration of the four-point calibration in the complex impedance plane for one sample (green diamond) for a defined measurement frequency. Calibration points are: calibration sample with lower conductivity than sample and without lift-off (blue circle) and with lift-off (blue triangle); calibration sample with higher conductivity than sample and without lift-off (red circle) and with lift-off (red triangle)

## Instrumentation

The hardware for use of high-frequency eddy current requires tools similar to a conventional eddy current system: A clean sinusoidal voltage needs to be generated and be supplied to the excitation coil. The measured voltage signal must be amplified and its amplitude and phase are then compared with those of the excitation signal. Specific circuits for stabilization, amplification, and electrical impedance matching of the excitation and measurement signals must be provided. For HFEC, the electronics must be able to operate properly at frequencies between 100 kHz and 100 MHz.

An important factor in HFEC testing is the coil system. Most applications use a minimum of two coils, one for excitation and one for measurement. They can be positioned next to each other or can be wound on each other. What is important is that the coils have maximum sensitivity in the frequency range of interest. Therefore the main resonance of the coils in comparison with the component which has to be analyzed should be higher than the highest frequency in the range. To reach that goal, coils with a low number of turns are used, sometimes without any core. These coils have in general a low sensitivity, but one has to find a compromise between desired sensitivity and response in that frequency range. The best suited coils for frequencies between 10 and 100 MHz are spiral coils with a very low number of turns. Figure 3 shows some examples of suitable coils.

Due to the use of these kinds of coils, special handling and calibration of the system is necessary. Additionally, it is advantageous to separate the preamplifiers of the driving and the outgoing electronic circuit of the electronic from the main board and bring them very near to the sensors. This minimizes the influence of the cable between coils and electronics on the signal (Jia et al. 2009).



**Fig. 3** Suitable coil-configuration for HFEC: spiral coils for frequencies higher than 10 MHz (left), real example (middle), and half transmission coils with two wound coils on a ferrite core (right)

## Typical Applications

### High-Frequency Eddy Current Techniques for Application to Carbon Composites

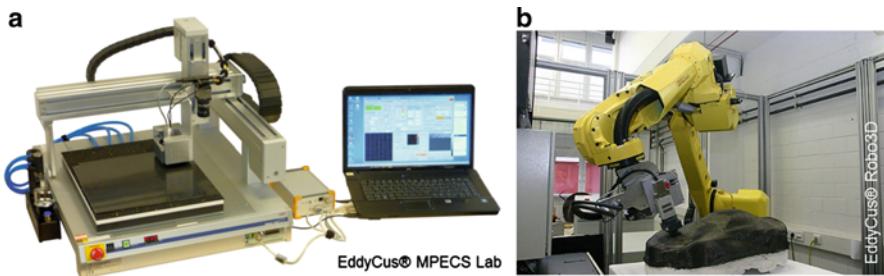
#### Basic Principles of HFEC for Carbon Composites

Carbon fiber materials are becoming increasingly important in many applications. Using the HFEC-technique, it is possible to detect structural and hidden defects such as missing carbon fiber bundles, missing fiber lanes, suspensions, fringes, missing sewing threads, and angle errors. Detection is independent of the material being flat or curved, or if the material exists as a woven raw material or in the resifined condition (CFRP). The penetration depth depends on the density of the material and the sensor configuration and can reach several millimeters or around 11 layers of carbon fibers.

For flat components, as they mostly are available as woven raw-material, a simple scan desk with minimum 2 axes, but ideally 3 axes can be used (Fig. 4, left). For curved components or for components with more complex geometries, a robot system is a better choice (Fig. 4, right).

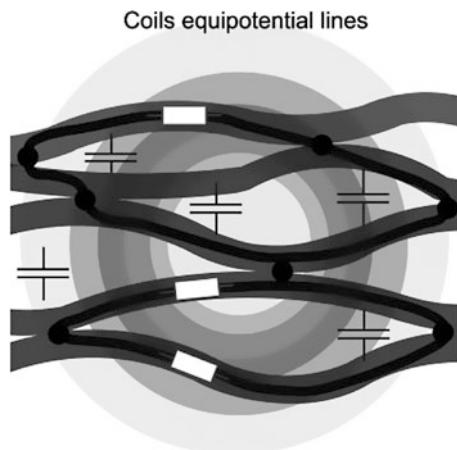
Unidirectional single-layered carbon fiber material has a conductivity of up to  $\sigma = 5 \cdot 10^6$  S/m in the longitudinal direction and  $\sigma = 1 \cdot 10^3$  S/m in the lateral direction. Through variation of fiber type, fiber orientation, stacking sequence, fiber/volume content, fiber density, and compaction around the carbon rovings, these values may differ. In addition to the conductivity of the CFRP, its permittivity is also important in RF testing. The displacement current in the material is proportional to permittivity, and it influences the measurement signal as well as the eddy current. In a dry carbon material, the permittivity is dominated by the fiber coating and the surrounding air. In a CFRP, the air is substituted by a polymer resin, so that the permittivity depends on type and processing quality of the resin. Figure 5 shows the main electrical effects of applying an alternating magnetic field to strongly anisotropic CFRP material that are influenced by three main parameters.

1. Fiber/Volume Ratio: The Fiber/Volume ratio determines the amount of conductive carbon fibers in a volume section and defines the average electrical



**Fig. 4** Two different HFEC-Scan-Systems: scan desk with three axes (left) and robot system (right)

**Fig. 5** Electric and dielectric behavior of carbon matrix composites



conductivity of the material, which needs to be taken into account for general measurement parameterization (frequency, penetration depth).

2. Electrical contact between fiber bundles: Depending on the structure (woven, crimped, non-crimped, etc.), chemical and structural conditions of the interfaces of filaments, and the consolidation density due to mechanical pressing, the electrical contact between neighboring bundles can vary. Identical materials can provide different degrees of eddy current propagation due to the quality of internal electrical connections during consolidation. The horizontal electrical connection between fiber bundles (in plane, parallel to the surface) is directly influencing the contrast of the RF image. The vertical connection in depth direction is influencing the interlaminar interfaces and thereby the penetration depth.
3. Capacitive effects and displacement current: In addition to the electrical contact between fiber bundles, the dielectric properties of the matrix material also influence the complex signal impedance.

These three parameters can all be analyzed by HFEC instruments. Medium-frequency eddy current (MFEC) inspection instruments are available in many configurations from various commercial suppliers. Typically, the instruments are

designed for manual handling of a single coil sensor or array probe with wheel tracker. The frequency range and parameterization of such standard devices is optimized for conventional NDE tasks such as crack detection or material identification for metallic specimens. Also, mechanical manipulators like XYZ axis scanners or robot-based manipulators are in commercial use to scan an EC probe over a three-dimensional surface (e.g., rotor blades). For low conductive carbon-based materials, a contrast enhancement at frequencies up to 50 MHz, and in some special cases up to 80 MHz, was observed. One reason for this is the increasing signal strength due to the Faraday law of induction, a second reason is the decrease in impedance of the capacitive structure of CFRP at higher frequencies.

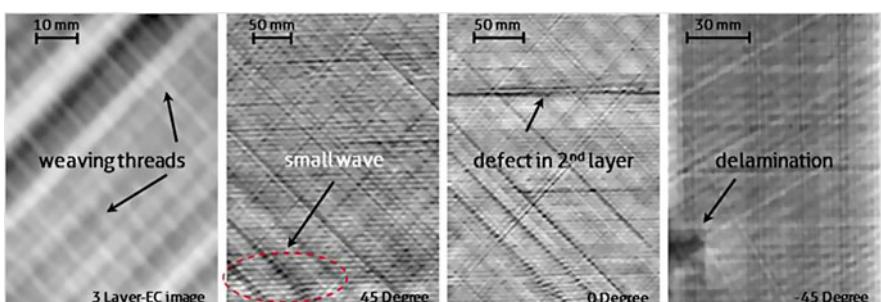
To perform eddy current measurements in the frequency range above 1 MHz, mechanical vibrations (lift off variations) and electromagnetic disturbance need to be controlled precisely. Also, electrical conductive dust in carbon contaminated environments can deteriorate the measurement results. To solve these practical problems, a robust setup needs to be used that is protected against dust and has an Electro Magnetic Safety architecture.

To generate HFEC Scans on Carbon fiber materials, a special half-transmission HFEC-sensors with directional properties is scanned over the surface of the samples. A typical half transmission probe configuration includes two 5 mm coils. Given the geometric distance between the sending and receiving coil, the sensitive field of view of the probe is elliptical and the observed spatial resolution is higher than the core diameter. Figure 6 demonstrates some typical scans with typical defects.

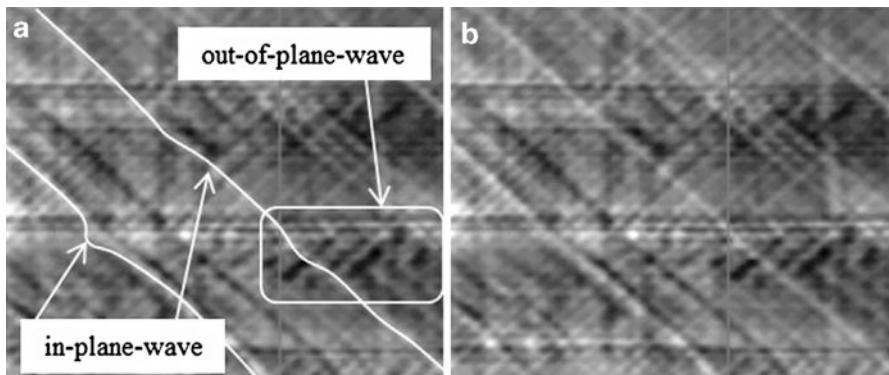
### Inspection of Texture and Waviness

The HFEC technique can be applied to dry, wet, and consolidated carbon-based materials. One major task of HFEC tests is the characterization of waves inside carbon fiber or consolidated materials. The image in Fig. 7 shows in-plane and out-of-plane waves perpendicular to each other.

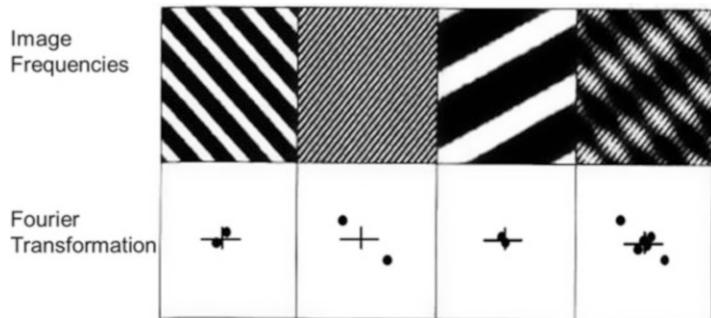
The In-plane wave can be seen directly due to the typical not straight forward orientation of the fiber bundle. An out-of-plane wave shows a modulation of the signal amplitude as gray value contrast. Due to local changes in fiber volume content, the signal amplitude changes depending on the out-of-plane wave position.



**Fig. 6** Typical HFEC-Scans on CFRP-components with a number of defects



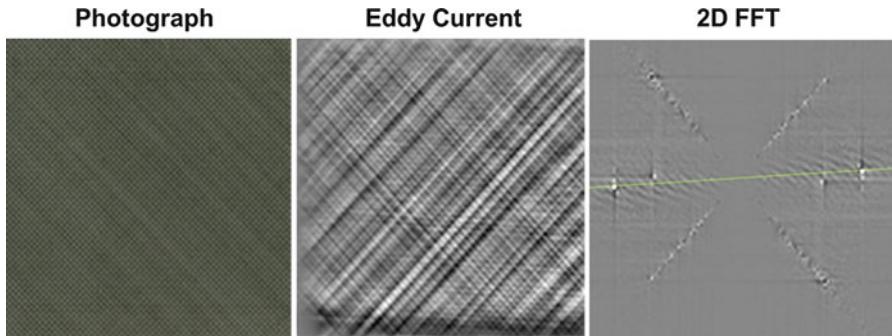
**Fig. 7** HFEC Images of CFRP plates with in-plane and out-of-plane waves (Size  $150 \times 150 \text{ mm} \times 5 \text{ mm}$ ). (a) indication of wave position and (b) raw data image



**Fig. 8** Theoretical approach of the 2D Fourier analysis of HFEC-Scans on material made of fibers: The Fourier transformation gives the angle of the points, the angle of the fibers, and the distance of the points to the middle point of the width of the fibers

In Fig. 8, the out-of-plane waves are indicated by black contrast change compared to the typical undulation of in-plane waves. Due to this different contrast mechanism, it follows that in-plane waves and out of plane waves require different algorithms for evaluation. For industrial application of HFEC imaging techniques, the textural information such as roving orientation, gap sizes, or waviness needs to be extracted by image processing. In particular, for multi-axial materials, a two-dimensional fast Fourier transformation (2D-FFT) shows high potential for automated fiber texture analyses.

In textured materials the image frequency is correlated with the periodic structure. In relation to the texture of CFRP fabrics, the image frequency correlates with gap size and fiber bundle size whereas the rotation of the image frequency maxima represents a layer orientation. Figure 9 (middle) shows a HFEC Image of a 3 axial non-crimp fabric with two horizontal layers on the back side. The corresponding 2D-FFT of Fig. 9 shows the three characteristic lines indicating the  $45^\circ$ ,  $0^\circ$ , and  $-45^\circ$



**Fig. 9** Example of 2D Fourier analysis on a sample of CFRP. A misalignment of the angle in one layer can be clearly detected. In the scan (middle) this property is difficult to detect

layer. Inside the 2D-FFT data in Fig. 9 (right), each line of points contains information for one specific CFRP orientation. By setting a filter that suppresses all points except the point bar from the layer and performing an inverse 2D-FFT, the textural image of one individual layer can be reconstructed.

An alternative method to analyze the texture of CFRP materials is the Polarlab-technique. In this method, a one-directional sensor rotates at a single point on the carbon composite component and generates HFEC-data for different angles of the sensor. The device is the so-called EddyCus® Polarlab (Fig. 10, left). The data generated can be displayed in a polar diagram (Fig. 10, right). Depending on the number, orientation, and amplitude of the peaks, very exact information of the fiber orientation of the different layers can be generated.

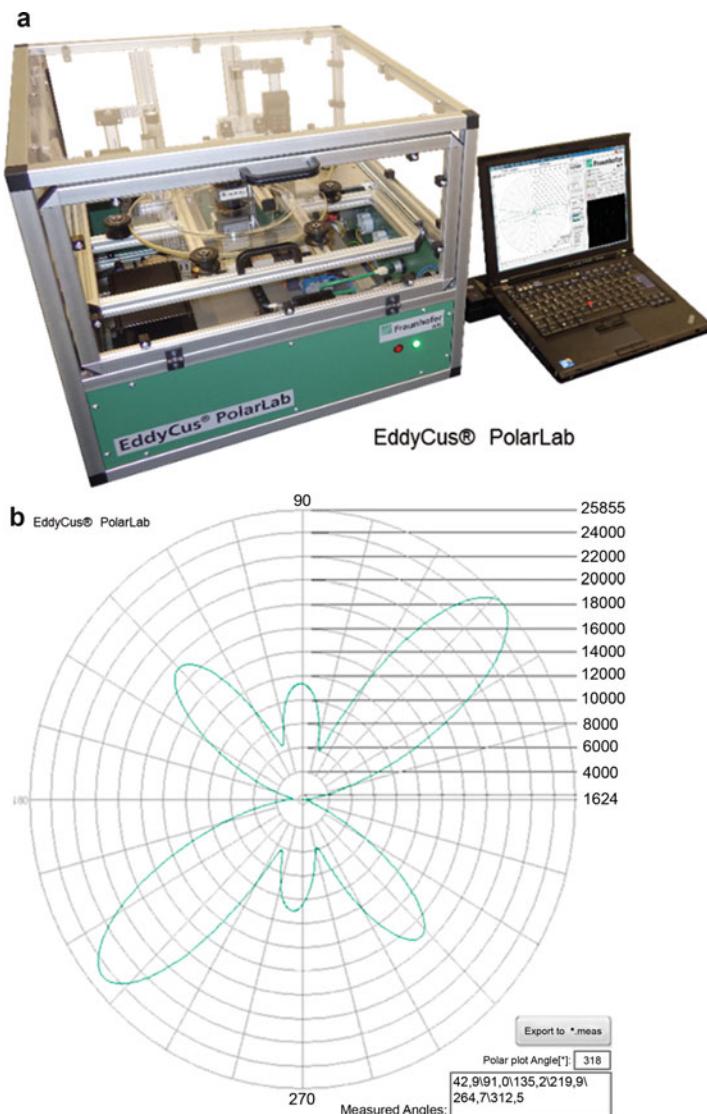
### Characterization of Defects According to Their Depth

The modulation of the signal amplitude as a function of the depth of an out-of-plane wave can be explained using Eqs. (1) and (2).

$$\delta = \sqrt{2/\omega\sigma\mu} \quad (1)$$

$$J(d) = J_s e^{-d\delta} \quad (2)$$

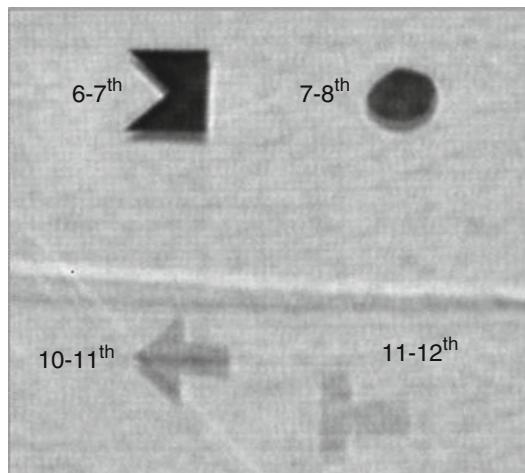
The eddy current density decays exponentially with increasing depth. For an isotropic material with good conductivity, the current density can be calculated by Eq. (2), where  $J_s$  is the Current density at the surface,  $d$  is the depth of the object, and  $\delta$  the penetration depth shown in Eq. (2). However, due to the strong anisotropy of CFRPs, this relation can be used only as a rough estimation. To validate the in-depth resolution for a specific CFRP material, a reference sample with insertions at different depths must be used. Figure 11 shows an HFEC Image of a CFRP plate with 14 layers of GV 300 U TFX (12 k) non-crimped fabrics with a grammage of  $317 \text{ g/m}^2$  for each layer and fiber volume content of 67%. During the laydown process, pieces of copper foil were inserted between different layers. The shape of the copper foil was used to identify the insertion depth using the reference after curing.



**Fig. 10** HFEC-system for detection of fiber orientation (left) and typical result on woven raw-material displayed on a polar diagram (right)

To obtain a high-contrast image, the phase angle is optimized until a maximum signal amplitude is obtained. Due to the complex structure of the materials, this optimum phase angle changes very strongly with increasing depth of information in the component, but also irregularly. In this special case, the copper insertion was seen up to the 12th layer. The phase angle where the amplitude reaches a maximum shows a linear dependency up to the 7th layer from the surface. For the deeper

**Fig. 11** HFEC Image at 8.12 MHz of CFRP with inserted Copper foils at different depths



regions, between the 7th and 14th layer the phased angle seems to be constant or decreases slightly. This behavior was not investigated yet but was reproduced in additional experiments.

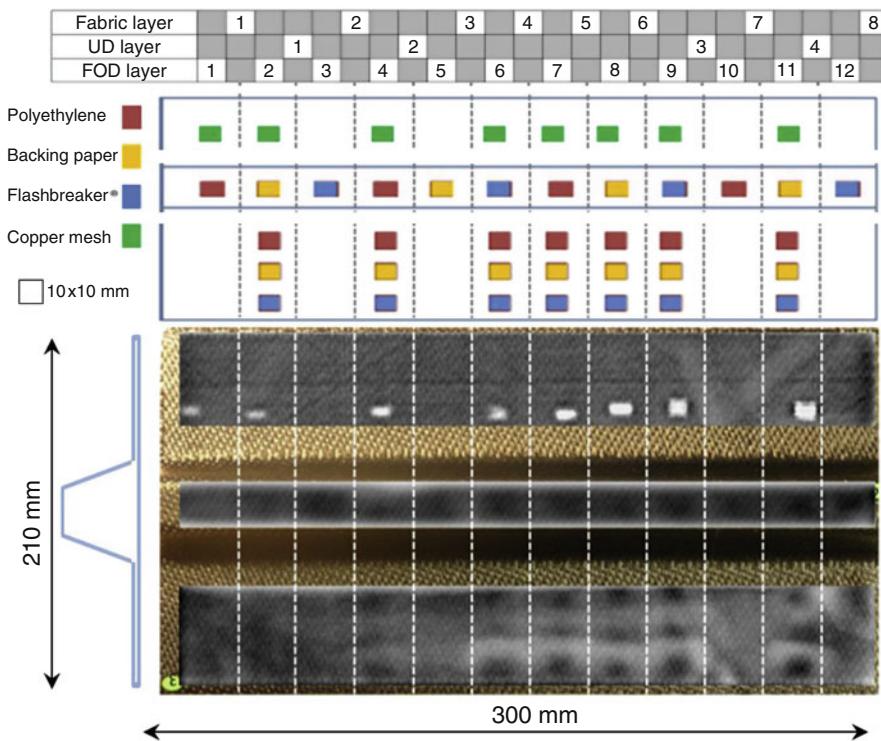
In addition to high conductivity, insertion materials like copper, lower or non-conductive materials could also be used and detected. A rectangular sandwich panel measuring  $300 \times 210$  mm with Foreign Object Damage (FOD) inclusions was made consisting of an  $\Omega$  stiffener of Rohacell 110 WF-HT covered with eight CFRP fabric layers (977-2A-37%-3kHTA-5H-280-1200) and four additional Uni Directional layers (977-2-34-12KHTS-196-T1-150) on top of the  $\Omega$  stiffener. Four types of FOD inclusions are inserted at different depths, each measuring  $10 \times 10$  mm:

1. Copper mesh (M21/67%/ECF73  $\varnothing$  V12/900 mm)
2. Flashbreaker<sup>®</sup> tape 0.18 mm thick
3. UD-backing paper (0.1 mm thick,  $\pm 130$  g/m<sup>2</sup>)
4. 0.13 mm thick Polyethylene prepreg release foil

The position of the insertion and the  $\Omega$  stiffener are illustrated in Fig. 12. The copper mesh was included as a reference and can still be detected under eight layers of CFRP fabric and four UD layers (in total 12 layers). The detection depth of the tape, polyethylene foil and paper are six fabric layers and three UD layers, respectively. On top of the  $\Omega$  structure, this is increased to seven CFRP fabric and three UD layers, respectively. The increased detection depth is probably due a reduced amount of air inclusions on the UD-layer on the top.

In another experiment, a CFRP plate with cut out wedges was produced. Figure 13a shows the position of the cut outs, Fig. 13b shows the corresponding HFEC image.

In the area of the cut out, the fiber orientation is disturbed but shaped CRFP parts often require a termination of hidden layers inside a multilayer stack (so-called ply



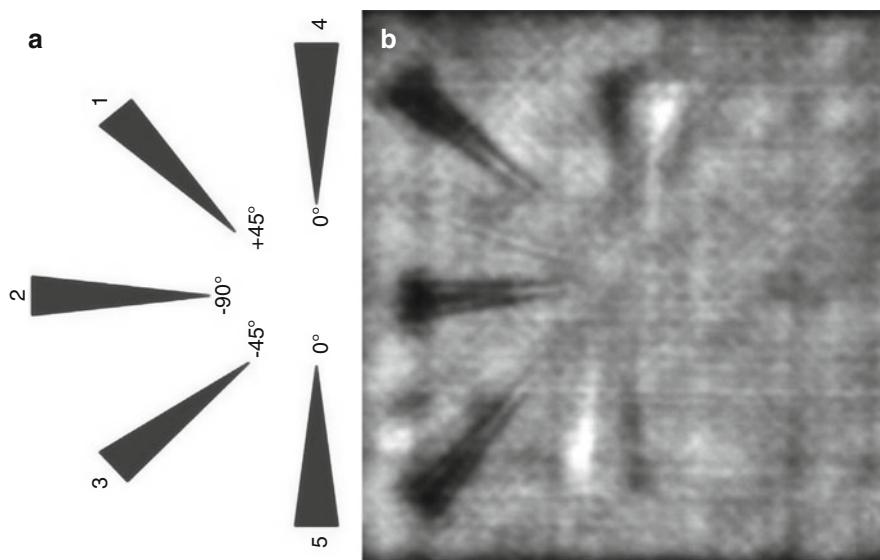
**Fig. 12** Schematic image FOD sample shows the type and position of different insertions with the corresponding RFEC image

drops). For quality assurance, the number of terminating layers must be determined (Abu-Nabah and Nagy 2007; Schulze et al. 2010; Lange and Mook 1994; Heuer et al. 2015).

## High-Frequency Eddy Current Techniques for Ceramic Composites

Ceramic-Matrix Composites (CMC) are damage tolerant ceramics that are increasingly established in technical structures, which work in harsh environment, at high temperatures, under corrosive attack, and high mechanical loads. CMC can be based on carbon fibers such as SiC (non-oxide CMC) or as oxide-fiber (oxide CMC) reinforcement. These composites are suitable for highly stressed components in mobile and stationary turbines.

Due to its known advantages for characterization of Carbon-Fiber-Materials, High-Frequency Eddy-Current techniques (HFEC) have also high potential to characterize CMC. Silicon Carbide (SiC) has a very low but present electrical conductivity. A mixture of inductive and capacitive effects as well as effects by displacement currents influences the measurement signal. Using these properties, it



**Fig. 13** HFEC Image of dry NCF plate with five UD Layers with cut out wedges in several orientations

is possible to characterize SiC-materials in solid state as well as in composites (CMC) with High-Frequency Eddy Current Scans, as the following figures demonstrate. Figure 14 shows a sample of SiSiCs: directed production of partially siliconized area (often a source for failure); not filled with Si: residual porosity and SiC.

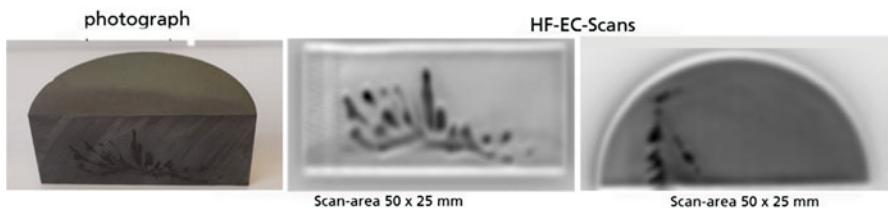
Figure 15 demonstrates a CMC-sample with long-fibers and irregular infiltration.

Figure 16 demonstrates that the HFEC works even in materials which consist of SiC fibers in a matrix of SiC. Figure 16 shows a photograph and an HFEC-Scan on a sample of this configuration with artificial defects in the forms of bore holes, notches, and flat-bottom holes.

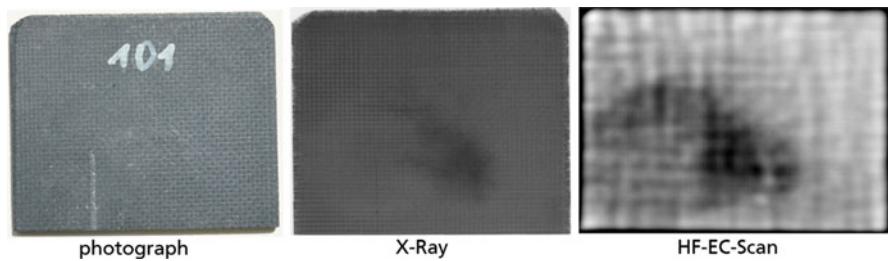
Due to much lower overall electrical conductivity, the HFEC technique has a higher penetration depth and a lower resolution in comparison to the application on CFRP (Langhof and Krenkel 2017; Kyrychenko et al. 2016).

### High-Frequency Eddy Current Techniques for Characterization of Materials with Low Electrical Conductivity

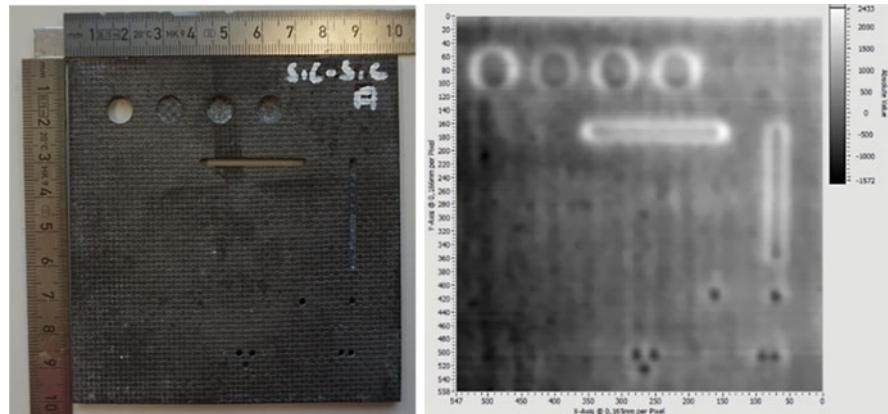
It is well known that it is possible to characterize material properties using eddy currents, when the material properties correlate with the electrical properties in the test materials. This mechanism works for the HFEC as well. Properties which can be characterized are mostly hardness and all properties which correlate with the structure, e.g., composition and amorphous structures. In some cases, attempts were made to measure quantitative values of the mechanical strain, attempts that were successful in some applications.



**Fig. 14** Photograph and HFEC-Scan on SiSiC-solid with partially siliconized area



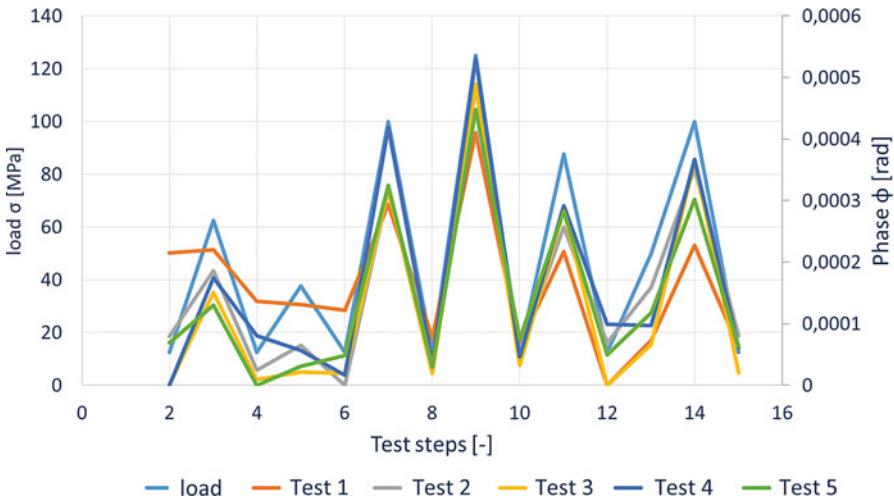
**Fig. 15** Photograph, Radiograph, and HFEC-Scan on CMC-sample with irregular infiltration (scan area is  $100 \times 100$  mm)



**Fig. 16** Photograph and HFEC-Scan on CMC-sample with artificial defects at a sample of SiC-fibers in a SiC matrix

### Measurement of Mechanical Load in CFRP Structures as Replacement to Strain Gauges

A useful and special application of HFEC methods is the measurement of mechanical load during the use of CFRP structures. Currently, the state of the art is the application of strain gauges to the component surface. However, carbon fibers change their electrical conductivity as a function of mechanical stress or strain according to the same principle as a strain gauge. Using the HFEC-technique, the



**Fig. 17** Results of one experiment on a bidirectional CFRP-sample (cross section  $40 \times 1 \text{ mm}^2$ ), delineated is the load profile (blue) and five repeated measurements

conductivity of carbon fibers can be measured without contact with the medium. The idea of this approach is to use the carbon fiber directly as a strain sensor. Therefore, it is possible for the first time to implement non-contact and long-life strain sensors, by elimination of age-sensitive mechanical bonding between the sensor and the component being monitored. The knowledge of the real load absorption and visualization of the load-bearing fibers enables fundamentally new possibilities in the development and use of CFRP components.

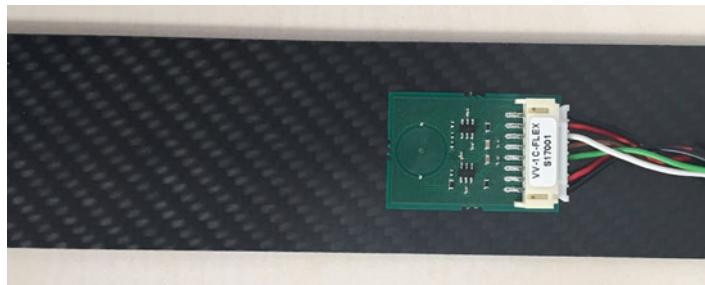
Analysis with flat spiral coils, which could be integrated into the structure of the CFRP or, alternatively, glued on the surface of components should result in a small, but basically valid measurement effect, whose best and stable variant is determined using the phase of the HFEC-signal (Fig. 17).

The measurement signal is very small and subject to variations due to component composition, component thickness, and width. Additionally the variations due to changes in the strain depend on the basic load: when the basic load is higher, the HFEC-signal due to strain variations decreases and that due to the relative signal variation decreases as well. This clearly indicates that frequent calibrations are needed as well as a sensor, which is, ideally attached to the component and measures the relative strain variations at one position of the component over a long time-period.

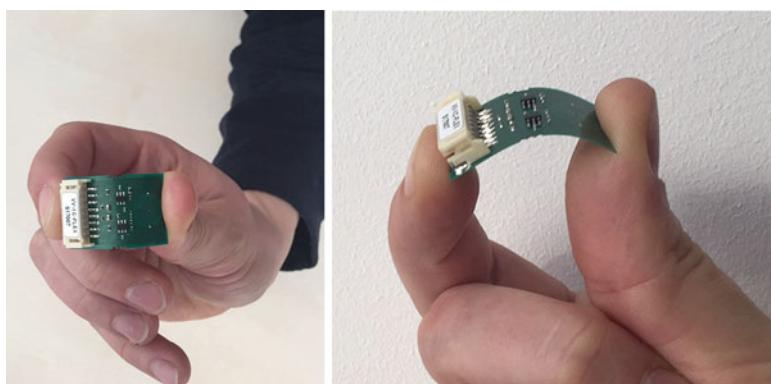
Therefore a Sensor consisting of a flat spiral HF-coil connected to a preamplifier could be used. Figures 18 and 19 show an example of such a sensor. All elements are applied on a flexible foil sensor, meaning that the sensor can be applied on curved surfaces as well.

### Characterization of Surface Layers by Sweeping Technology

One defined measurement mode is to permit the sweeping of numerous measurement frequencies in a large frequency range. Using this so-called sweep-mode,



**Fig. 18** Photo of a sensor patch with a simple cable connection glued on a CFRP test sample

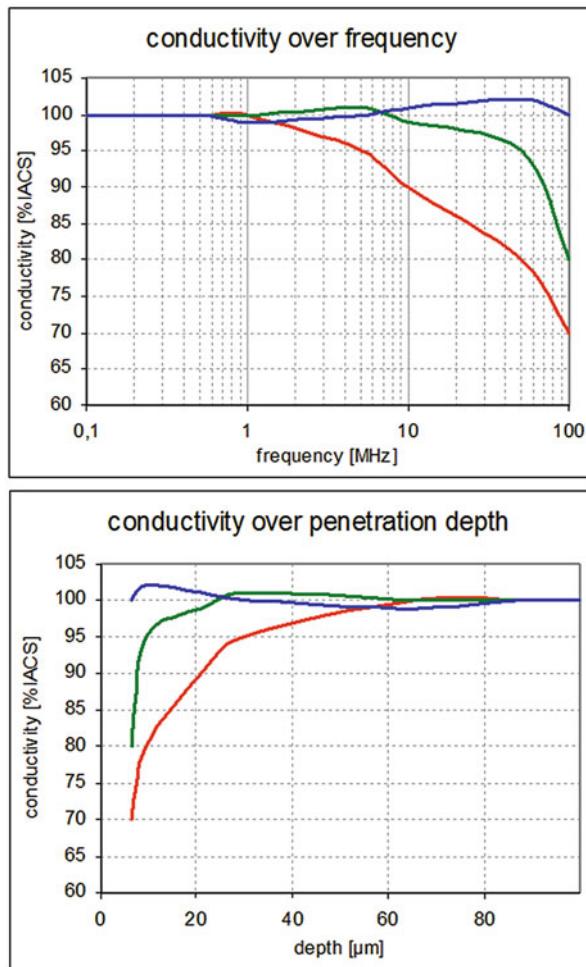


**Fig. 19** Photographs of a sensor patch; Demonstration of the flexibility of the patch

frequency dependent profiles of the eddy current measurement can be obtained. Therefore, the behavior of the eddy current parameters is visible over a defined frequency range and as a consequence, over different penetration depths. This makes the sweep-mode a good tool to analyze and monitor behaviors and effects on coating systems. By using the so-called four-point-calibration, first frequency profiles of each sample can be generated, which are defined as the apparent eddy current conductivity (AECC) as a function of the frequency. With the knowledge of the electrical conductivity, it is possible to calculate AECC profiles over the depth using the information of the penetration depth at the defined measurement frequencies (Fig. 20).

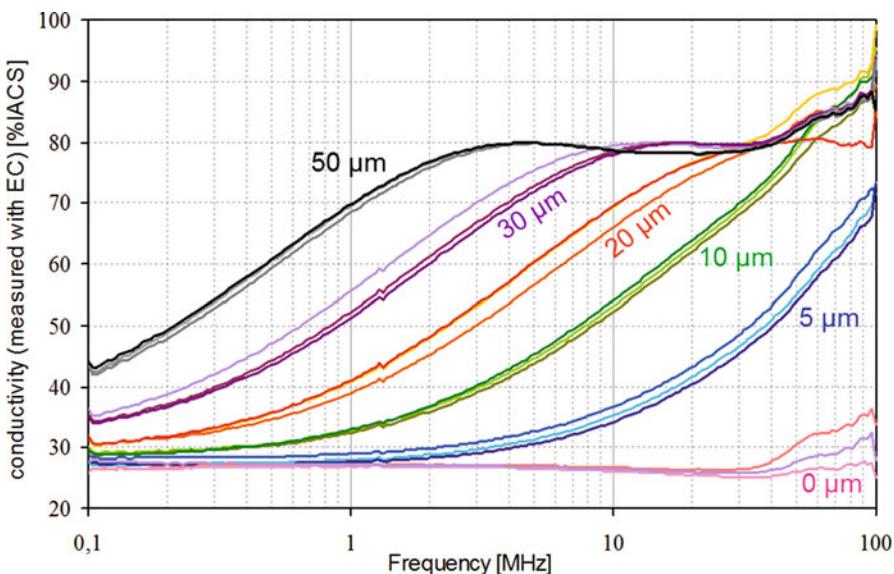
The measured AECC-profiles using HFEC depends strongly on the thickness of the layer and additional on the difference in conductivity between the substrate and the layer. That is why, both of these influencing factors has to be obtained. To get information about potentials and limitations of this relationship, sample sets with different single and double layer systems were designed. The samples with single layers consist of substrates with low, medium, and high electrical conductivity that are coated with layers of low, medium, and high electrically conductive layers

**Fig. 20** Example of conductivity profiles as a function of frequency (top) and accordingly, as a function of the depth (down)

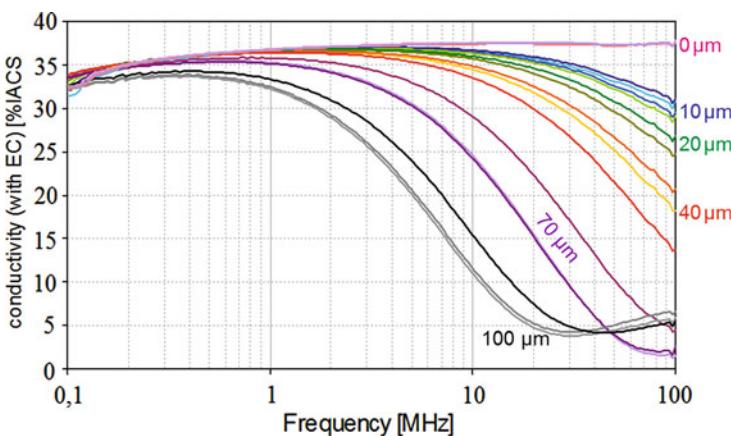


having different layer thicknesses. In addition, some samples are produced as double-layer-systems. For example, using a coating system with a silver layer on a substrate of copper, where both, substrate and layer, have a very similar high electrical conductivity, the developed technique is capable of exploring the limits of small differences in the conductivity between coating and substrate.

The following examples demonstrate the use of AECC frequency profiles of different single layer systems. Figure 21 shows the AECC of a sample set with substrates of brass and coatings of silver. The substrate has a medium conductivity, while the layer has a very high conductivity. therefore, a medium difference between the conductivity of substrate and layer was performed. The uncoated substrates have a constant electrical conductivity at around 30% IACS. The apparent conductivity increases, when more of the electromagnetic field is located in the silver layer. The thinner the layer is, the greater the increase in the signals at higher frequencies.

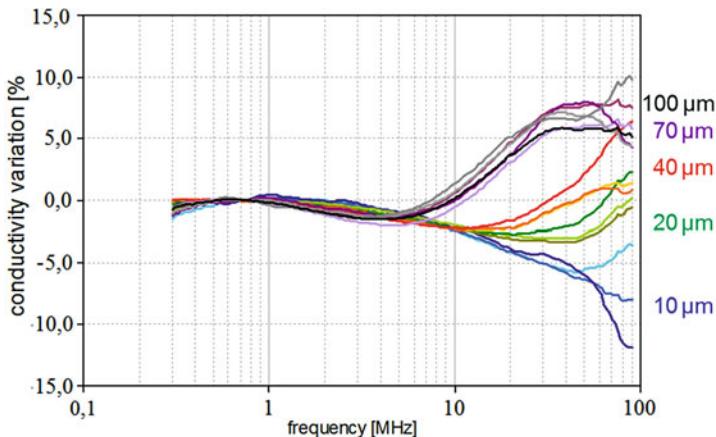


**Fig. 21** Frequency profiles of the apparent eddy current conductivity (AECC) for a sample set with substrates of brass and coatings of silver



**Fig. 22** Frequency profiles of the AECC for a sample set with substrates of brass and coatings of chromium

Figure 22 shows the apparent conductivity of a sample set with substrates of brass and coatings of chromium. This combination demonstrates the reverse case: the substrate has a medium conductivity, while the layer has a very low conductivity. Therefore again there is a medium difference between the conductivity of substrate and layer. The uncoated substrates have again a constant conductivity at around 30% IACS. The conductivity decreases, when the electromagnetic field is located more in



**Fig. 23** Frequency profiles of the AECC for a sample set with substrates of titanium and coatings of chromium

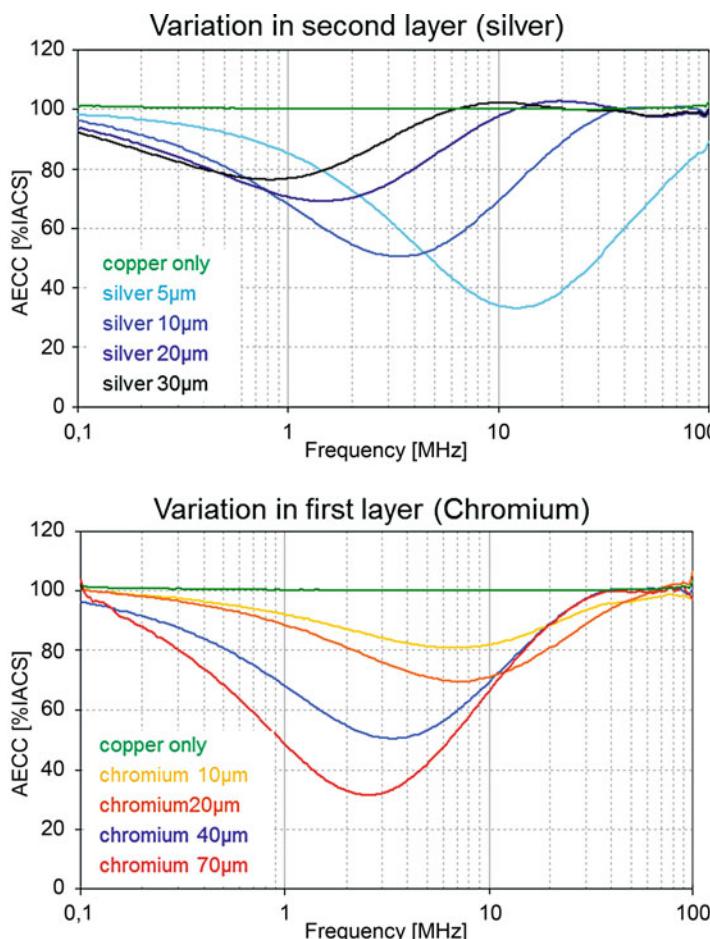
the chromium layer. The thinner the layer is, the greater the decrease of the signal at higher frequencies.

Figure 23 shows the apparent conductivity of a sample set with substrates of titanium and coatings of chromium. In this case, the substrate and layer both have a low electrical conductivity. Here a very small difference between the conductivity of substrate and layer was created. The variations of the AECC in the different layer thicknesses are very low. The electrical conductivity seems to increase at the higher measurement frequencies, while the layer thickness increases.

More interesting results can be generated, when double layers with different electrical conductivity are used. Figure 24 shows the AECC of such a sample set with a double-layer structure. The substrate is copper; the first coating is of chromium and the second coating of silver.

### Characterization of Surface Treated Materials for Aero Engine Components

On the other side of the spectrum of conductivities, attempts to determine the stress profile in the surface of shot peened Nickel-based Superalloys (e.g., Inconel 718) or titanium alloy (Ti-Al-6-4) were conducted. These typical aviation engine alloys were surface-treated to induce near-surface compressive strains. To calculate the remaining operation time for those critical (aviation) engine components, a quantitative nondestructive determination of near-surface strain gradients was needed. Due to the very low electrical conductivity of these alloys, the HFEC was very well suited. It could be demonstrated in different experiments that it is possible to determine a profile of the variations of the HFEC at different depths of the material's surface layer. But these variations were a combination of different material properties in the shot peened area such as strain, stress, and cold work. It was not possible to separate the information from the residual stress. However, it could be

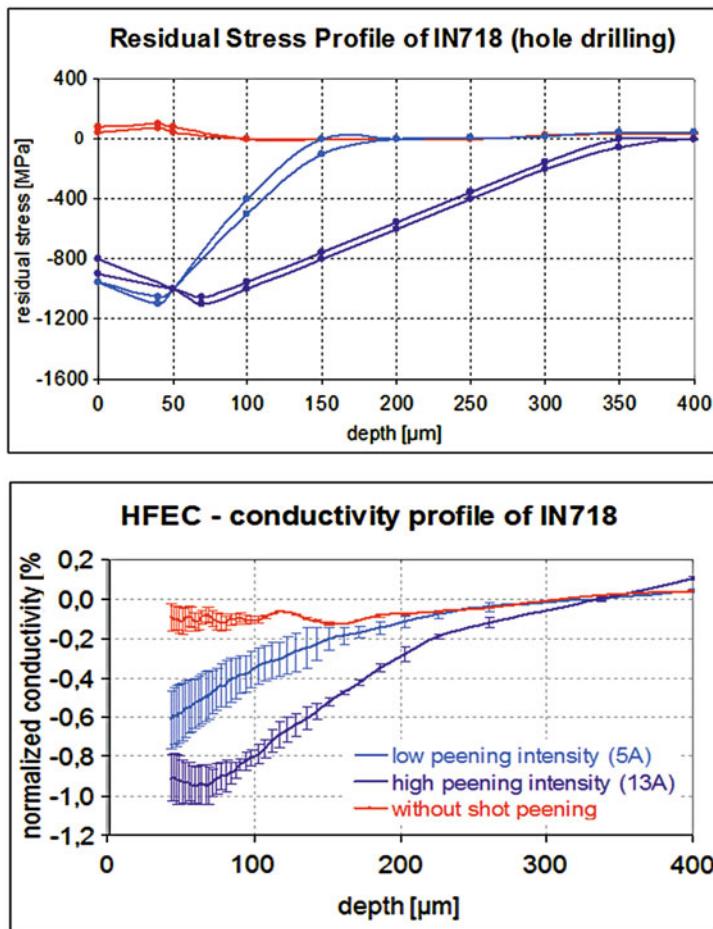


**Fig. 24** Frequency profiles of the AECC of a double-layer sample set; left: chromium layer is a constant 40 µm thick and the silver layer varies in thickness; right: silver layer is constant at a thickness of 10 µm and the chromium layer varies

demonstrated that it is possible to generate depth profiles of material properties in the surface by sweeping different frequencies using the HFEC. Figure 25 shows residual stress profiles of three precipitation hardened samples of IN718 obtained by a hole drilling method, and the profiles were determined by HFEC.

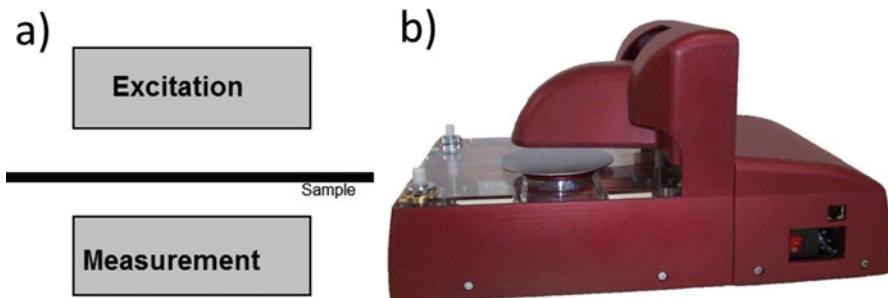
### Characterization of Film Thicknesses in the Nanometer Range

A third topic is the characterization of thin films in the nanometer range regarding thickness. Thin films are successfully used in many different ways as functional layers in all fields of science, technology, and medicine. The coatings are produced using different methods such as evaporation, sputtering, or chemical vapor deposition and have different properties such as adhesion, specific electric resistance, and



**Fig. 25** Left: residual stress profiles of three precipitation hardened samples of IN718 in three different states of shot peening intensity obtained by a hole drilling method, right: surface profiles determined by HFEC on the same samples

diffusion coefficient. For all applications, the analysis of the coating thickness plays an important role, because most of the operating and performance characteristics are a function of the coating thickness. Here in particular are highlighted in-line-methods, working contact-less acting, very fast and in a vacuum. For coatings with thicknesses in the nanometer range and the lower micrometer range, HFEC is very well suited. With a special sensor and application in the transmission mode, the method is very well adapted for in-line inspection. The coils in the transmission mode are separately arranged at a fixed distance from each other. The sample must be thin and flat and must be aligned between the sensors, which in most application of thin film technology is the case. Figure 26a illustrates a schematic of this application.

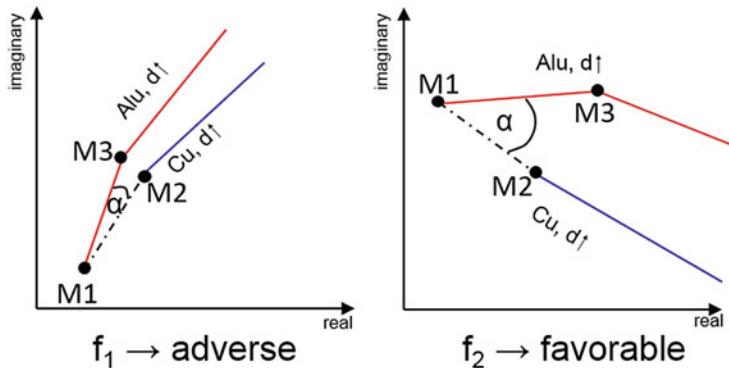


**Fig. 26** (a) Schematic illustration of the eddy current method for flat samples by use of transmission mode; (b) EddyCus TF® eddy current desktop system operating in the transmission mode

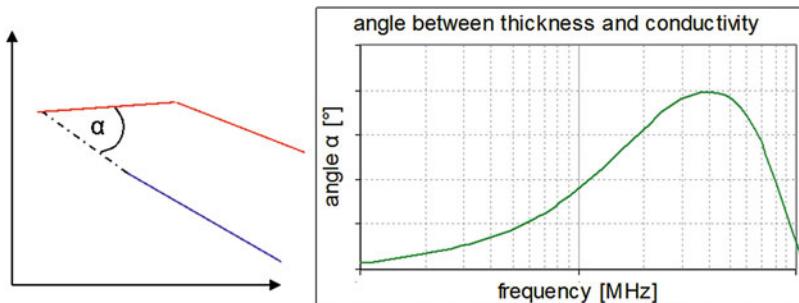
The advantage is low influence of misalignment of the sample between the sensors on the test result, making this very practical for in-line inspections. Additionally, it is possible to achieve a larger distance between sensor and sample using the transmission mode. In the solar industry, a large distance between sensor and sample is often required, especially when massive solar cells that are prone to bending loads pass through the coating line. Using the System in Fig. 26b, the sensor-sensor distance can be increased up to 60 mm, while still maintaining a high thickness resolution.

For very small thicknesses, it is especially important to choose the optimal measurement frequency to establish a good separation between data from sample thickness variations and from conductivity variations. This is particularly important when process-related temperature changes may alter the conductivity of the sample. Depending on the measuring frequency, both material properties can be separated from each other in varying degrees; Fig. 27 illustrates this relationship. The schematic illustrates the points in the complex impedance plane for measurements of thin metal sheets of different thicknesses and conductivities. Point M1 is obtained from the measurement on a thin sheet with conductivity  $\sigma_1$  (e.g., aluminum) and thickness  $d_1$ . The data at point M2 is generated by a measurement on a same material sheet (conductivity  $\sigma_2 = \sigma_1$ ), but higher layer thickness  $d_2 > d_1$ . In contrast, the data at point M3 results from the measurement on a sheet of the primary thickness of  $d_3 = d_1$ , but with a material of higher conductivity  $\sigma_3 > \sigma_1$  (e.g., copper). The left image shows this correlation at an unfavorable measurement frequency. Based on natural variations, the measurement points M2 and M3 are difficult to separate from each other. They are located very close to each other, because the impedance, locus of conductivity, and thickness are almost superimposed. The image on the right illustrates the complex impedance plane for the same samples at a suitable measurement frequency. In this case, points M2 and M3 are well separated from each other.

A dimension for separability of both properties is the opening angle  $\alpha$ , located within the starting point M1 and points M2 and M3. The larger the angle  $\alpha$ , the better conductivity and coating thickness can be separated in the measurement signal, and small film thicknesses in the nanometer range can be obtained with high resolutions even at higher conductivity variations.



**Fig. 27** Schematic illustration of the complex impedance plane with an adverse measurement frequency  $f_1$ , which produces poor separation of conductivity and thickness, and results in a small angle  $\alpha$  (left); a favorable measurement frequency  $f_2$  with a large opening angle  $\alpha$  (right)



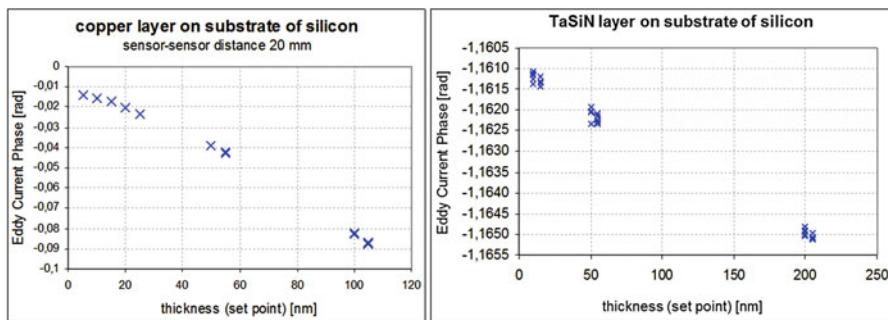
**Fig. 28** Determination of optimal measurement frequency for high conductivity by calculating the opening angle  $\alpha$  from measured data using thin aluminum and copper sheets with different thicknesses

To determine the optimal measurement frequency for higher conductivity of around 50MS/m, a series of experiments using thin metal sheets of aluminum and copper with different wall thicknesses were performed. The opening angle  $\alpha$  was calculated for each data set at the different measurement frequencies. The results are illustrated in Fig. 28. The maximum opening angle  $\alpha$  was determined and can be used as the optimal measurement frequency for this conductivity range. Each conductivity range has its own optimum measuring frequency, whereas the optimum measuring frequency increases with decreasing film conductivity.

For quantitative measurements of the film thickness, a calibration is needed. Therefore samples of the same material composition and the same geometry but different layer thickness are needed. The minimum number of samples in the calibration set is two: one sample with thickness in the minimum thickness area to be measured and one in the higher thickness range. The accuracy of the measurement can be increased using more samples for calibration which are located more in the

**Table 1** Accuracy of the tested samples at different sensor-to-sensor distances, obtained in the HFEC transmission mode

Coating material	Coating thickness range	Sensor-to-Sensor distance 20 mm	Sensor-to-Sensor distance 40 mm	Sensor-to-Sensor distance 60 mm
<b>Single layer</b>				
Copper	5–100 nm	±1 nm	±1.5 nm	±2 nm
TaSiN	10–200 nm	±20 nm	±50 nm	±100 nm
<b>Multi Layer</b>				
Silver	100–600 nm	±10 nm	±15 nm	±17 nm
Aluminum	100–600 nm	±21 nm	±36 nm	±35 nm
Total	100–600 nm	±54 nm	±55 nm	±104 nm



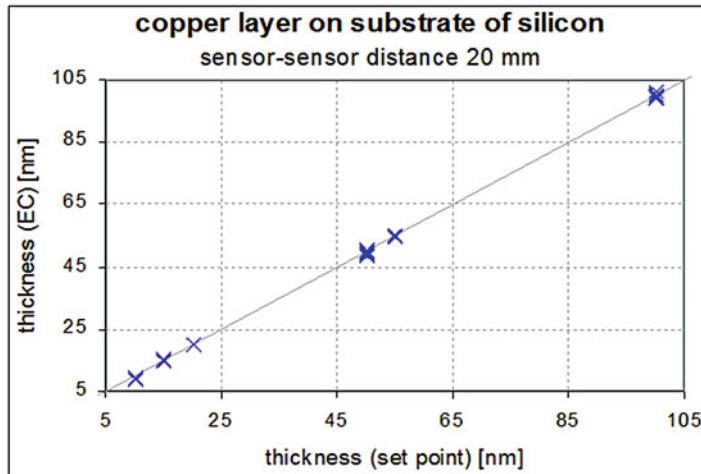
**Fig. 29** The phase of the HFEC signal as a function of the set point of the coating thickness; left: results of thickness measurements on sample of copper layer on silicone; right: results of thickness measurements on sample of a TaSiN layer on silicone

middle range of the thickness range. Especially when the thickness ranges, which has to be measured, is larger, more than two samples for calibration are advisable, due to the fact that curves in the complex impedance plane are not linear. Only in a small range of the measurement range they can be supposed as linear.

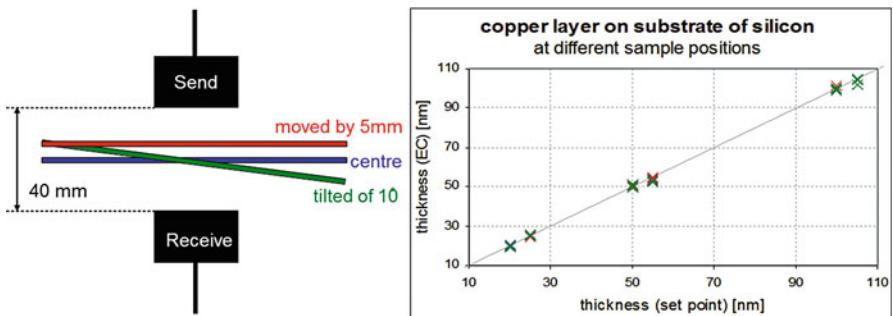
Analysis on different coatings of Copper and TaSiN (thallium silicon nitride) as well as combinations of these showed very precise and reproducible quantitative results partially in the lower nanometer range, as the following figures and the table demonstrate (Abu-Nabah et al. 2009; Hillmann et al. 2008, 2010; Blodget and Nagy 2004) (Table 1) (Figs. 29, 30, 31, and 32).

## High-Frequency Eddy Current Techniques of Polymers and Dielectrics

When the HFEC system is applied on dielectric materials, significantly higher displacement currents are generated by the time-varying magnetic field than in



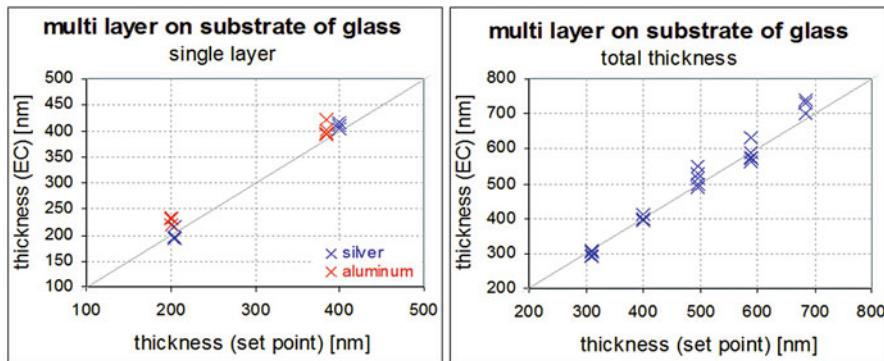
**Fig. 30** Coating thickness as a function of the given value of the film thickness established by HFEC on samples of copper layer on silicone



**Fig. 31** Left: Schematic illustration of the sample positions tested in the measuring gap; right: results of thickness measurement at different sample positions

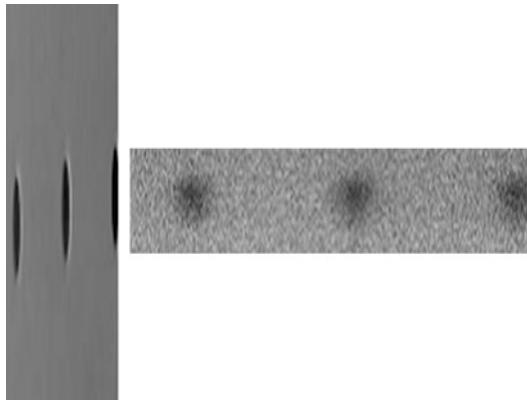
eddy current measurement set ups with lower frequencies (in the kHz-range). That effect can be used to measure permittivity-related properties with HFEC, especially on electrically low-conducting or insulating materials. Figure 33 demonstrates as an example an HFEC-Scan of a dielectric Polyoxymethylen (POM) sample.

Experiments showed, quantitative measurements of complex permittivity with HFEC are also possible using an appropriate calibration. Measurement accuracy is comparable to commercial capacitive dielectric analyzers. A very practical application of this technique is the monitoring of the curing process of epoxy resin, for example, for CFRP-materials. These measurements were performed using



**Fig. 32** Coating thicknesses obtained by eddy current techniques as a function of the film thickness given value, measured on sample set 3 (multiple layers). Left: measurements of total thickness (all layers). Right: measurements of single layers

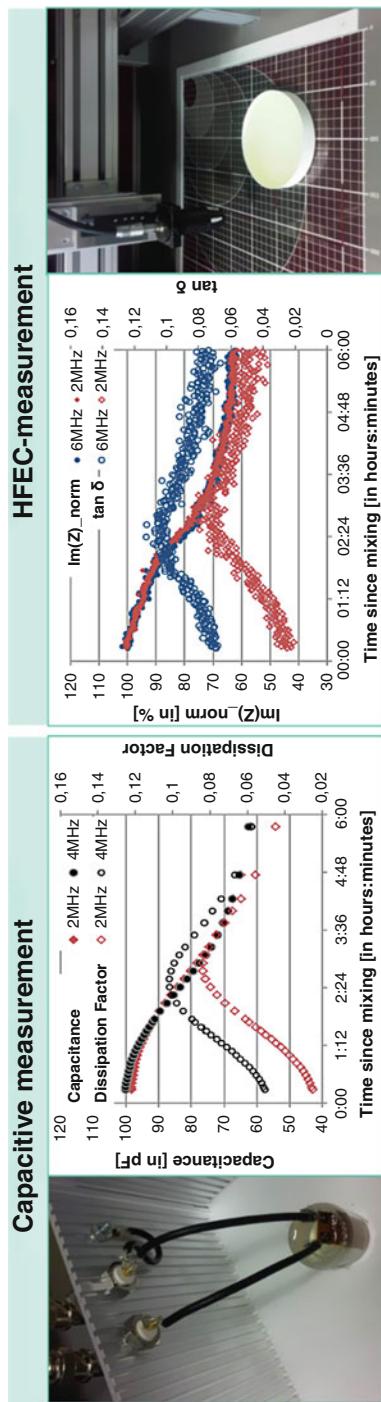
**Fig. 33** Left: photograph of a sample of Polyoxymethylen (POM) with bore holes, right: HFEC-Scan on the same sample



destructive capacitive measurements and can be replaced with the nondestructive HFEC-technique. The results are completely comparable, as Fig. 34 shows.

## Summary

The High Frequency Eddy Current (HFEC) technique is a special eddy current method, which works in the higher frequency range around 100 kHz to 100 MHz. It needs special hardware which is optimized for this frequency range and especially specific coils, which are sensitive in that range. It can be used for defect detection or material characterization of all kinds of materials with very low electrical conductivity including nonconducting materials. In these applications, it is possible to reach higher sensitivities and resolution with optimized penetration depths compared to electromagnetic NDE-techniques working with lower frequencies.



**Fig. 34** Measurement set up for characterization of the curing process of epoxy resin with HFFC-measurement (left) and capacitive measurement (right: the sensor remains in the resin); the results of the two methods are superimposable (Gäbler et al. 2015a, b).

## References

- Abu-Nabah B, Nagy PB (2007) Lift-off effect in high-frequency eddy current conductivity spectroscopy. *NDT&E Int* 40:555–565
- Abu-Nabah BA, Yu F, Hassan WT, Blodgett MP, Nagy PB (2009) Eddy current residual stress profiling in surface-treated engine alloys. *Nondestruct Test Eval* 24:209
- Blodgett MP, Nagy PB (2004) Eddy current assessment of near-surface residual stress in shot-peened nickel-base superalloys. *J Nondestruct Eval* 23:107–123
- Gäbler S et al (2015a) Quantitatively analyzing dielectrical properties of resins and mapping permittivity variations in CFRP with high-frequency Eddy current device technology. In: 41st annual review of progress in quantitative nondestructive evaluation, AIP conference proceeding, vol 1650, pp 336–344. AIP Publishing LLC. <https://doi.org/10.1063/1.4914628>. 978-0-7354-1292-7
- Gäbler S et al (2015b) Measuring and imaging permittivity of insulators using high-frequency Eddy-current devices. *IEEE transducers on instrumentation and measurement*, 0018-945, IEEE
- Heuer H, Hillmann S, Röllig M, Schulze M, Wolter K-J (2009) Thin film characterization using high frequency eddy current spectroscopy. In: 9th IEEE conference on nanotechnology, NANO 2009, pp 165–168. Conference on nanotechnology (NANO)'9, Genoa
- Heuer H, Schulze M, Pooch M, Gäbler S (2015) Process monitoring for resins, carbon fiber fabrics, preforms and consolidation by HF radio wave techniques. In: 14th international symposium on nondestructive characterization of materials (NDCM 2015) June 2226, Marina Del Rey
- Hillmann S, Heuer H, Baron H-U, Bamberg J, Meyendorf N (2008) Near-surface residual stress-profiling with high frequency eddy current conductivity measurement. In: 35th annual review of progress in quantitative NDE, QNDE. Chicago 20–25 July
- Hillmann S, Dayoub R, Röllig M, Meyendorf N, Heuer H (2010) High-frequency Eddy current conductivity measurement for near-surface layer characterization. In: Knopp J et al (eds) Electromagnetic nondestructive evaluation (XIII). IOS Press BV, Netherlands, pp 79–86
- Jia Y, Heuer H, Hillmann S, Meyendorf N (2009) Stray capacitances of an air-cored Eddy current sensor. *Sensor Transducers J* 111(12):25–37
- Kyrychenko M et al (2016) NDT investigations on C/SiC samples from different manufacturing steps. In: Proceedings of the 8th international symposium on NDT in aerospace, 3–5 Nov
- Lange R, Mook G (1994) Structural analysis of CFRP using eddy current methods. *NDT&E Int* 27(5):241–248
- Langhof N, Krenkel W (2017) Ceramic matrix composites for high performance friction applications. In: Proceedings of the IV advanced ceramics and applications conference, Jan 2017, pp 13–28. [https://doi.org/10.2991/978-94-6239-213-7\\_2](https://doi.org/10.2991/978-94-6239-213-7_2)
- Schulze M-H, Heuer H, Küttner M, Meyendorf N (2010) High-resolution eddy current sensor system for quality assessment of carbon fiber materials. *Microsyst Technol* 16(5):791–797



# Eddy Current Tomography

21

Antonello Tamburrino and Guglielmo Rubinacci

## Contents

Introduction .....	758
Monotonicity Principle for Eddy Current Imaging in the Large Skin-Depth Regime .....	761
Imaging Method .....	762
The Sign Index .....	764
Examples of Application .....	765
Monotonicity Principle for Eddy Currents in Other Settings .....	769
Small Skin-Depth Regime .....	769
Pulsed Eddy Current Imaging: Time Constants .....	772
Summary .....	775
References .....	777

### Abstract

In eddy current tomography, the conductivity profile of conductive materials is reconstructed through the inversion of eddy current data (ECT). The state of the art of imaging methods in ECT data inversion is represented by iterative methods, the drawbacks of which are their high computational cost and the risk of becoming trapped in false solutions (local minima). In this chapter, we discuss the “Monotonicity Principle Method,” a fast non-iterative approach recently developed for elliptic problems (such as electrical resistance tomography) and then extended to parabolic

---

A. Tamburrino (✉)

Department of Electrical and Information Engineering, Università degli Studi di Cassino e del Lazio Meridionale, Cassino, Italy

Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA

e-mail: [antonello.tamburrino@unicas.it](mailto:antonello.tamburrino@unicas.it)

G. Rubinacci

Department of Electrical Engineering and Information Technology, Università degli Studi di Napoli Federico II, Napoli, Italy

e-mail: [rubinacci@unina.it](mailto:rubinacci@unina.it)

problems (such as eddy current tomography) and hyperbolic problems (such as microwave tomography).

This chapter discusses the main features of the Monotonicity Principle in eddy current testing. Specifically, section “[Monotonicity Principle for Eddy Current Imaging in the Large Skin-Depth Regime](#)” discusses the Monotonicity Principle for eddy current testing in the “large” skin-depth regime, then section “[Imaging Method](#)” introduces the related imaging method (Monotonicity Principle Imaging Method, MPIM), and section “[Monotonicity Principle for Eddy Currents in Other Settings](#)” introduces other setting where Monotonicity Principle holds and MPIM can be applied.

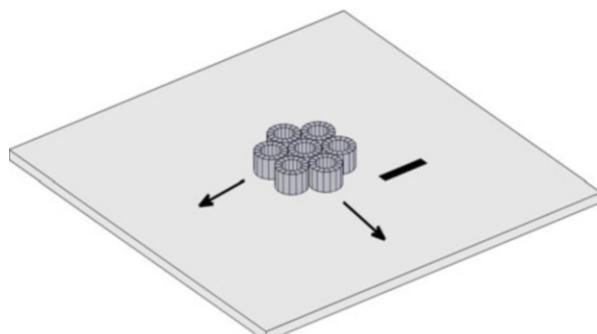
## Introduction

In eddy current tomography, a given time-varying current flowing in the excitation coils induces eddy currents in the conductive specimen under inspection. The electrical resistivity of the specimen affects the induced eddy current density and, consequently, the magnetic flux density measured by a set of sensors. Both coils and sensors are external to the conductors.

Starting from the measurement data, the conductivity profile of the test specimen is reconstructed through the solution of an inverse problem. Hereafter we assume the measured data being the impedance matrix (self and mutual impedances) between the coils of an array (see Fig. 1).

The eddy current problem is modeled by a parabolic partial differential equation (PDE). While the solution of the forward problem, i.e., the calculation of the magnetic flux density and/or pickup coil voltages/impedances, is well-established, the inverse problem is ill-posed and nonlinear (see Isakov 1993; Yamamoto 1997 for mathematical issues) and hence remains a challenge. As a matter of fact, the measurement (magnetic flux density and/or pickup coil voltages/impedances) is a nonlinear function of the electrical resistivity in the specimen. Moreover, the sensitivity of the measurement w.r.t. the electrical resistivity distribution in the material is very poor. Other critical issues are the sensitivity to position and tilting of the probes and skin-effect phenomenon.

**Fig. 1** The specimen under test together with a probe made up of an array of seven coils and a defect. (Reprinted from Tamburrino et al. 2010 with permission from IOP Publishing Ltd.)



The state of the art of inversion methods for eddy current tomography mainly involves iterative methods based on quadratic and higher-order approximations of the forward operator, deterministic and stochastic algorithms, pre-calculated databases, statistical methods, total variation regularization, level set regularization, etc. (see Yusa et al. 2002, 2003; Abu-Nabah and Nagy 2006; Stolzenburg et al. 2003; Premel and Baussard 2002; Norton and Bowler 1993; Bowler 2002; Luong and Santosa 1998; Badics et al. 1998; Li et al. 2004; Yin et al. 2005; Nair and Rose 1990; Pirani et al. 2008; Dorn and Lesselier 2009; Qi-Nian 2000; Calvano et al. 2010; Soleimani and Lionheart 2006; Lionheart et al. 2003, and references therein). The main drawbacks of iterative methods are their high computational cost (at least one forward problem has to be solved at each iteration) and the risk of becoming trapped in false solutions. Opposed to iterative methods, there are the so-called non-iterative methods. Non-iterative imaging methods are very well suited to evaluate the support of anomalies, as they provide a simple test for evaluating whether or not a point in an otherwise homogeneous body is part of the anomaly. The test is usually very cheap from the computational viewpoint (iterations are not necessary), and, moreover, it is a local test, i.e., it can be performed on a specific point of the space independently of the other points. Therefore, the number of tests to be performed is equal to the number of points considered in sampling the domain thus making non-iterative methods suitable for real-time applications.

To date, only a few non-iterative imaging methods are known. Colton and Kirsch introduced the first non-iterative approach named linear sampling method (LSM) (Colton and Kirsch 1996) for wave-propagation phenomena. The factorization method (FM) was proposed by Kirsch (1998) for wave-propagation phenomena and extended by Hanke and Brühl to electrical resistance tomography (ERT) (Hanke and Brühl 2003). MUSIC (MULTiple SIgnal Classification) algorithm is well known in signal processing. Devaney (2000) pointed out that it could also be used in imaging. Ammari (Ammari et al. 2005) developed a MUSIC algorithm for locating small inclusions buried in a half-space, given the scattered amplitude data. This method is able to detect multiple inclusions at the same time but limited to the case that inclusions are small and well separated. Eventually, the Monotonicity Principle Method (MPM) was introduced by Tamburri and Rubinacci (2002) and applied to electrical resistance tomography (ERT). In this method a simple test determines if a given small test anomaly is contained or not in the unknown anomaly. MPM was first introduced in the area of ERT (Tamburri and Rubinacci 2002; Tamburri 2006 (static problems, Laplace equation)) and then extended to eddy current testing in both small skin-depth regime (Tamburri et al. 2010) and large skin-depth regime (Tamburri and Rubinacci 2006; Tamburri 2006), with experimental validations in Tamburri et al. (2012) and (time domain) pulsed eddy current testing in Su et al. (2017b). MPM was also extended to (wave propagation) inverse scattering problems (Tamburri et al. 2015). MPM provides upper and lower bounds for the unknown object in the case of a finite number of measurements (limited aperture data) (Tamburri and Rubinacci 2002). In Harrach and Ullrich (2013), it was proven that MPM provides the exact shape of the anomalies in the ideal case of an infinite number of

measurements (full-aperture data), under the hypothesis that each connected component of the anomalies is contractible.

Among these methods, MPM is a topic of active research and has been shown to be promising by many investigators. In Garde and Staboulis (2017), Garde et al. have formulated a Monotonicity-based shape reconstruction scheme and regularized against noise and modeling error. Further, two reconstruction algorithms are formulated and validated by numerical examples with both simulated complete electrode model (CEM) data and experimental measurement data. In Garde (2018), the author compared the full nonlinear and a linearized version of MPM, proving that they give similar accuracy at similar computational cost, in the unit disk domain. In medical imaging area, a novel bimodal imaging technique combining both breast microwave radar (BMR) and EIT methods is presented to form a resistivity distribution map of a breast region that can be used to assess the presence of malignant lesions (Flores-Tapia and Pistorius 2010; Flores-Tapia et al. 2011). This technique uses a priori information obtained from BMR images to estimate the location of the dense breast regions, and then, the Monotonicity of the impedance matrix is used to reconstruct a profile of the tissue distribution in the breast region. In Aykroyd et al. (2005), Aykroyd et al. investigated the performances of MPM on ERT problems. Specifically, a modified algorithm integrating Bayesian modeling and MCMC (Markov chain Monte Carlo) estimation was proposed for configurations with a limited number of electrodes and low noise level. In Wallinger et al. (2009), Wallinger et al. combined MPM and Gauss-Newton-based iterative algorithms for electrical capacitance tomography of a two-phase scenarios consisting of gas/oil mixtures. MPM was used to provide the initial guess for the Gauss-Newton method which, in turns, was used to refine the boundaries of the anomalies. In Harrach et al. (2015), Harrach et al. developed a Monotonicity-based imaging method by combining frequency-difference EIT with ultrasound modulated EIT. The advances of this method are that it does not require any numerical simulation or forward model or even the knowledge of the imaging domain. This method is validated with 2D and 3D numerical models but relies on the assumption that an ultrasound wave could be perfectly focused in a small test region.

Interesting to notice is that Monotonicity Principle provides a rather general framework which can be customized for specific needs. Other than for imaging, MP can be used for sizing and material characterization (Su et al. 2017a, 2018). Sizing refers to the estimate of some key size of an anomaly, regardless the exact shape. This problem finds a large interest in industrial applications where one needs a rough estimate about the size of the anomaly, rather than its exact shape. MP has also been used as homogenization technique (Maffucci et al. 2016) where one is interested in finding an homogeneous material which can “reasonable” replace the original (non-homogeneous) material. Among the more interesting customization of MP, we mention the possibility of providing bounds to the solution of the inverse problems (Tamburrino and Rubinacci 2002; Tamburrino et al. 2016). This is a unique feature of MP among the inversion methods. Moreover, MP can be found in other different settings, such as static problems (Laplace equation) or wave-propagation problems, as aforementioned.

## Monotonicity Principle for Eddy Current Imaging in the Large Skin-Depth Regime

In this section we briefly summarize the Monotonicity Principle as from Tamburri and Rubinacci (2006). We assume an array of coils (such as that in Fig. 1) as the probe and the impedance matrix between all the pairs of coils in the array as the measured data.

Let  $D$  be the conducting domain and let  $\eta$  be the electrical resistivity. We assume that  $0 < \bar{\eta}_L \leq \eta(\mathbf{r}) \leq \bar{\eta}_H < +\infty$  being  $\bar{\eta}_L$  and  $\bar{\eta}_H$  two constants. Hereafter the impedance matrix is termed as  $\mathbf{Z}_\eta$  whereas its real and imaginary parts as  $\mathbf{R}_\eta$  and  $j\omega\mathbf{L}_\eta$ , respectively ( $\mathbf{Z}_\eta = \mathbf{R}_\eta + j\omega\mathbf{L}_\eta$ ), being  $\omega$  the angular frequency (the  $e^{j\omega t}$  time-dependence is assumed). In the low-frequency limit, the impedance matrix admits the following expansion (Tamburri and Rubinacci 2006):

$$\mathbf{Z}_\eta = \mathbf{R}_0 + j\omega\mathbf{L}_0 + \sum_{n=1}^{+\infty} \omega^{2n} \mathbf{P}_\eta^{(2n)} + j \sum_{n=1}^{+\infty} \omega^{2n+1} \mathbf{P}_\eta^{(2n+1)} \quad (1)$$

where  $\mathbf{R}_0$  and  $\mathbf{L}_0$  are the static resistance and inductance matrices in the free space and the  $\mathbf{P}_\eta^{(m)}$  are proper real and symmetric matrices.

The Monotonicity Principle states that the second-order moment  $\mathbf{P}_\eta^{(2)}$  is monotonic w.r.t. the electrical resistivity, i.e.,

$$\eta_1(\mathbf{r}) \leq \eta_2(\mathbf{r}) \quad \forall \mathbf{r} \in D \Rightarrow \mathbf{P}_{\eta_1}^{(2)} \geq \mathbf{P}_{\eta_2}^{(2)}. \quad (2)$$

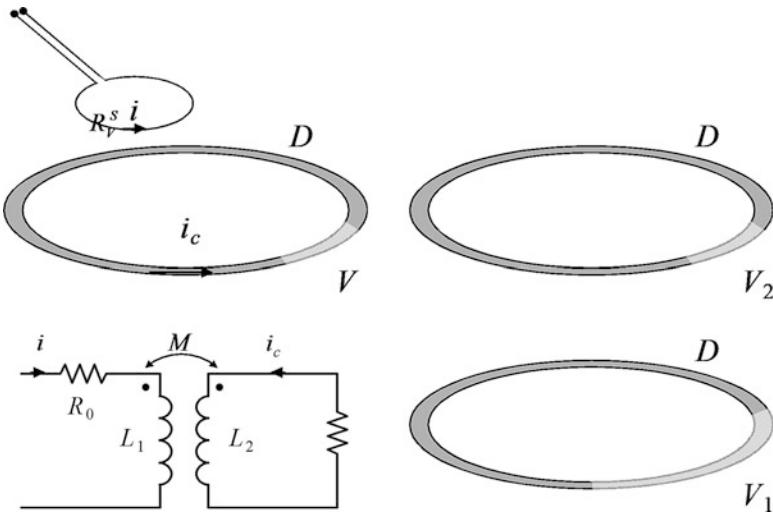
In (2)  $\mathbf{P}_{\eta_1}^{(2)} \geq \mathbf{P}_{\eta_2}^{(2)}$  means that matrix  $\mathbf{P}_{\eta_1}^{(2)} \geq \mathbf{P}_{\eta_2}^{(2)}$  is positive semi-definite, i.e., all its eigenvalues are nonnegative. When the fourth-order term and higher are negligible in (1), as in the case of large skin-depth regime (low-frequency regime), it follows that  $\mathbf{R}_\eta = \mathbf{R}_0 + \omega^2 \mathbf{P}_\eta^{(2)}$  and, therefore, (2) translates as

$$\eta_1(\mathbf{r}) \leq \eta_2(\mathbf{r}) \quad \forall \mathbf{r} \in D \Rightarrow \mathbf{R}_{\eta_1} \geq \mathbf{R}_{\eta_2}. \quad (3)$$

Monotonicity (3) is a very general property. It can be specialized to two-phase materials, which is the common case when one is interested in detecting/sizing/imaging homogeneous anomalies in an otherwise homogeneous background. Specifically, let  $\eta_B$  be the electrical resistivity of the background and let  $\eta_A$  be the resistivity of an anomaly. Assuming that  $\eta_B < \eta_A$  as it is for common practical applications, we have that (3) can be cast as

$$V_1 \subseteq V_2 \subseteq D \Rightarrow \mathbf{R}_{V_1} \geq \mathbf{R}_{V_2} \quad (4)$$

where  $V_1$  and  $V_2$  are two possible anomalies and  $\mathbf{R}_{V_1}$  and  $\mathbf{R}_{V_2}$  are the corresponding resistance matrices. In other words, if we enlarge the anomaly (from  $V_1$  to  $V_2$  which includes  $V_1$ ), then the resistance matrix measured at the probe inputs “increases” in some sense rigorously stated by the Monotonicity Principle. Equation (4) is the basis for the imaging algorithm.



**Fig. 2** Left: a simple configuration where a single excitation coil is used to probe a quasi-filamentary conductor (gray) having an equivalent resistance equal to  $R_V^S$ , together with its equivalent circuit. Right: two possible “anomalies” such that  $V_2 \subseteq V_1$  and  $R_{V_1}^S \geq R_{V_2}^S$ . (Reprinted from Tamburrino et al. 2012 with permission of Elsevier Ltd.)

In order to better grasp the concept of Monotonicity let us consider a simple problem consisting of a probe made up of one exciting coil and one quasi-filamentary conductor (see Fig. 2). The system is coupled magnetically (we are in the magneto-quasistatic limit), and it is equivalent to two coupled conductors (a transformer) with a current generator injecting a current on the “primary” and a proper resistor on the “secondary.” The equivalent resistance on the “primary” side is:

$$R_V^{Eq} = R_0 + \frac{R_V^S (M/L_2)^2}{1 + (R_V^S/\omega L_2)^2}, \quad (5)$$

where  $R_0$  is the resistance of the exciting coil;  $L_1$ ,  $L_2$ , and  $M$  are the self and mutual inductance of the coupled inductors; and  $R_V^S$  is the resistance of the quasi-filamentary conductor constituting the secondary of the “transformer.”

For small values of  $\omega$ , we have  $R_V^{Eq} = R_0 + \frac{\omega^2 M^2}{R_V^S}$ . Thus  $V_2 \subseteq V_1 \subseteq D \Rightarrow R_{V_1}^S \geq R_{V_2}^S \Rightarrow R_{V_1}^{Eq} \leq R_{V_2}^{Eq}$ , in line with (4).

## Imaging Method

Equation (4) represents the basis for developing a Monotonicity-based inversion method. To this purpose, let  $\mathbf{R}_V$  be the real part of the impedance matrix referring to an unknown anomaly  $V$  in  $D$ , and let  $\mathbf{R}_k$  be that related to a test domain  $\Omega_k$ . The test

domain is a prescribed anomaly of given geometry, position, and electrical resistivity equal to  $\eta_A$ .

From property (4), it follows that

$$\Omega_k \subseteq V \subseteq D \Rightarrow \mathbf{R}_k \geq \mathbf{R}_V, \quad (6)$$

therefore, if  $\mathbf{R}_k - \mathbf{R}_V$  is not positive semi-definite, then  $\Omega_k$  cannot be contained in  $V$ . Similarly, from (4) it follows that

$$V \subseteq \Omega_k \subseteq D \Rightarrow \mathbf{R}_k \leq \mathbf{R}_V, \quad (7)$$

therefore, if  $\mathbf{R}_k - \mathbf{R}_V$  is not negative semi-definite, then  $\Omega_k$  does not contain  $V$ .

We highlight that  $\mathbf{R}_k$ , which is related to the test domain  $\Omega_k$ , can be precomputed and stored once for all. On the other hand,  $\mathbf{R}_V$  corresponds to the unknown anomaly  $V$ , and it is experimentally measured.

From (6) and (7), two possible reconstruction strategies follow:

- Take as reconstructed domain the union of those  $\Omega_k$ 's such that  $\mathbf{R}_k \geq \mathbf{R}_V$ ;
- Take as reconstructed domain the intersection of those  $\Omega_k$ 's such that  $\mathbf{R}_k \leq \mathbf{R}_V$

In order to perform the Monotonicity tests (i.e., checking conditions  $\mathbf{R}_k \geq \mathbf{R}_V$  and  $\mathbf{R}_k \leq \mathbf{R}_V$ ), we need to compute the eigenvalues of the difference matrix  $\mathbf{R}_k - \mathbf{R}_V$ . In all practical cases, the measurements are corrupted by noise: only a noisy version  $\tilde{\mathbf{R}}_V$  of the measured matrix is available. In line with Harrach and Ullrich (2015) and Tamburrino et al. (2016), we assume to know the noise level  $\delta$  affecting the measurements, i.e.,  $\|\tilde{\mathbf{R}}_V - \mathbf{R}_V\| \leq \delta$ . From  $\|\tilde{\mathbf{R}}_V - \mathbf{R}_V\| \leq \delta$ , the two following inequalities hold:

$$\begin{aligned} \mathbf{R}_V &\geq \tilde{\mathbf{R}}_V - \delta \mathbf{I} \\ \mathbf{R}_V &\leq \tilde{\mathbf{R}}_V + \delta \mathbf{I} \end{aligned}$$

being  $\mathbf{I}$  the identity matrix. As a consequence (6) and (7) translate as

$$\begin{aligned} \Omega_k \subseteq V \subseteq D \Rightarrow \mathbf{R}_k &\geq \tilde{\mathbf{R}}_V - \delta \mathbf{I} \\ V \subseteq \Omega_k \subseteq D \Rightarrow \mathbf{R}_k &\leq \tilde{\mathbf{R}}_V + \delta \mathbf{I} \end{aligned}$$

and the corresponding imaging algorithms are:

- Take as reconstructed domain the union of those  $\Omega_k$ 's such that

$$\mathbf{R}_k \geq \tilde{\mathbf{R}}_V - \delta \mathbf{I}; \quad (8)$$

- Take as reconstructed domain the intersection of those  $\Omega_k$ 's such that

$$\mathbf{R}_k \leq \tilde{\mathbf{R}}_V + \delta \mathbf{I}. \quad (9)$$

Criterion (8) guarantees that if a test domain is contained into the anomaly, then it will not be discarded in the presence of noise. Similarly criterion (9) guarantees that if a test domain contains the anomaly, then it will not be discarded in the presence of noise. Hereafter we denote with  $V_U$  and  $V_I$  the reconstructions following from the *union rule* (8) and from the *intersection rule* (9), respectively.

It is worth noting that if the unknown domain  $V$  can be perfectly represented as union of test domains, then (Tamburrino et al. 2016):

$$V \subseteq V_U \quad (10)$$

Similarly, if the unknown domain  $V$  can be perfectly represented as intersection of test domains, then:

$$V_I \subseteq V. \quad (11)$$

Relationships (10) and (11) are important because they provide upper and lower bound to  $V$  even in the presence of noise.

The proposed technique is fast and powerful since it involves a very simple test on the eigenvalues of a suitable difference matrix. The size of this matrix is relatively small: equal to the number of coils of the excitation array that, in practical cases, is of the order of few dozens.

## The Sign Index

Another reconstruction strategy, which can be used when the noise level is unknown, is based on the so-called sign index  $s_k$  related to test anomaly  $\Omega_k$  (Tamburrino and Rubinacci 2006):

$$S_k = \frac{\sum_j \lambda_j (\mathbf{R}_k - \tilde{\mathbf{R}}_V)}{\sum_j |\lambda_j (\mathbf{R}_k - \tilde{\mathbf{R}}_V)|}, \quad (12)$$

where  $\lambda_j(\mathbf{R}_k - \tilde{\mathbf{R}}_V)$  is the  $j$ -th eigenvalue of the matrix  $\mathbf{R}_k - \tilde{\mathbf{R}}_V$ . The sign index in (12) “measures” how much the matrix  $\mathbf{R}_k - \tilde{\mathbf{R}}_V$  is positive semi-definite. It assumes values in the range  $[-1, +1]$ . Specifically  $s_k = +1 (-1)$  if  $\mathbf{R}_k - \tilde{\mathbf{R}}_V$  is positive (negative) semi-definite; otherwise  $|s_k|$  is strictly less than 1. As aforementioned, each  $s_k$  can be related to test anomaly  $\Omega_k$  and a function of the space can be easily built as follows:

$$S(\mathbf{r}) = S_k \text{ if } \mathbf{r} \in \Omega_k,$$

where we have assumed that the  $\Omega_k$  are not overlapped. Generalization to overlapped test domain can be easily carried out by using an approach similar to Tamburrino et al. (2010). We found that the plot of  $1/[1 - S(\mathbf{r})]$  provides a qualitative reconstruction of the unknown anomaly  $V$  (Tamburrino and Rubinacci 2006). We refer to Tamburrino et al. (2010) for extensions in the use of the sign index.

## Examples of Application

Many examples are available in Tamburrino and Rubinacci (2006) and Tamburrino et al. (2012). Here we present few results from the experimental tests carried out in Tamburrino et al. (2012).

The experimental tests we describe in this section refer to an array made up of two coils mounted in a fully automated imaging system comprising a scanning system, an LCR meter, and a PC controlling the whole acquisition and inversion process (see Fig. 3).

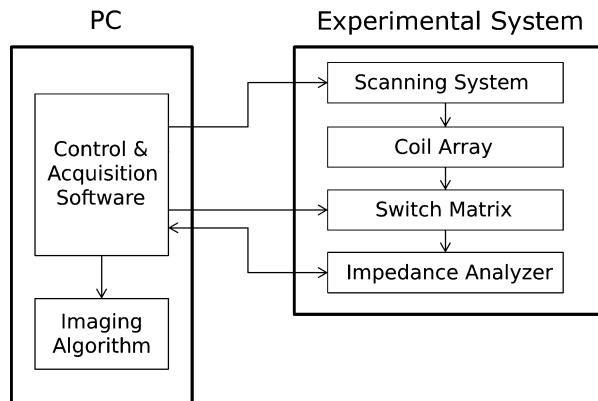
The specimen under test is a printed circuit board (PCB, thickness of the dielectric of 1.6 mm) presenting copper islands of different sizes and shapes to be imaged. The elementary (test) domains are  $5 \times 5$  mm (nonoverlapped) copper islands onto a regular 2D Cartesian grid. The unknown domains are a union of these elementary ( $5 \times 5$  mm) cells. We highlight that the resistance matrix for a test anomaly has been experimentally measured on a single  $5 \times 5$  mm copper island. Thanks to the translational invariance of the problem, this measurement on a single test domain also makes it possible to obtain the resistance matrix for any other test anomaly in the regular grid.

The sensor, which has been designed through extensive numerical simulations, is made up of two pancake coils (see Fig. 4). The first coil (internal diameter = 5 mm, external diameter = 10.5 mm, height = 6.5 mm, number of turns = 700) contains internally the second coil (internal diameter = 1 mm, external diameter = 4 mm, height = 3 mm, number of turns = 180).

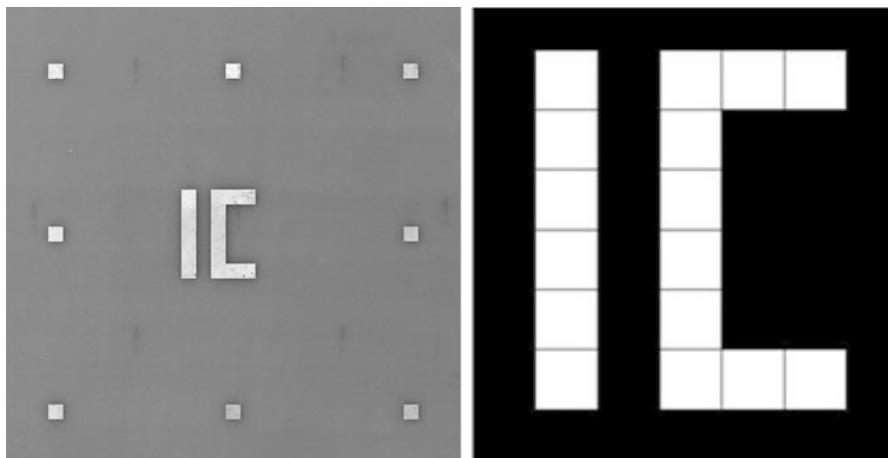
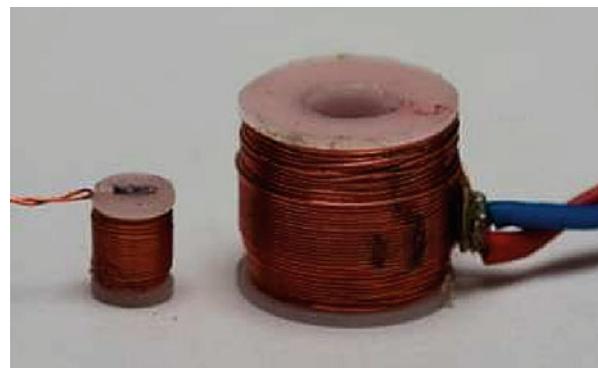
The excitation frequency is 30 kHz, chosen to satisfy the condition that the fourth- and higher-order terms in (1) are negligible (the skin-depth is about one order of magnitude larger than the thickness of the conductive layer of the specimen, and the magnetic diffusion time is about one order of magnitude smaller than twice the period of the applied time-harmonic waveform). At this frequency the magnitude of the noise level is smaller than  $20 \text{ m}\Omega$ .

The benchmark is subdivided into pixels of dimensions  $5 \times 5$  mm, and the sensor is moved over it while keeping its barycenter on the center of each region pixel. The

**Fig. 3** Block diagram of the measurement system



**Fig. 4** The two coils composing the array. The smaller coil is inserted into the larger one. (Reprinted from Tamburrino et al. 2012 with permission of Elsevier Ltd.)



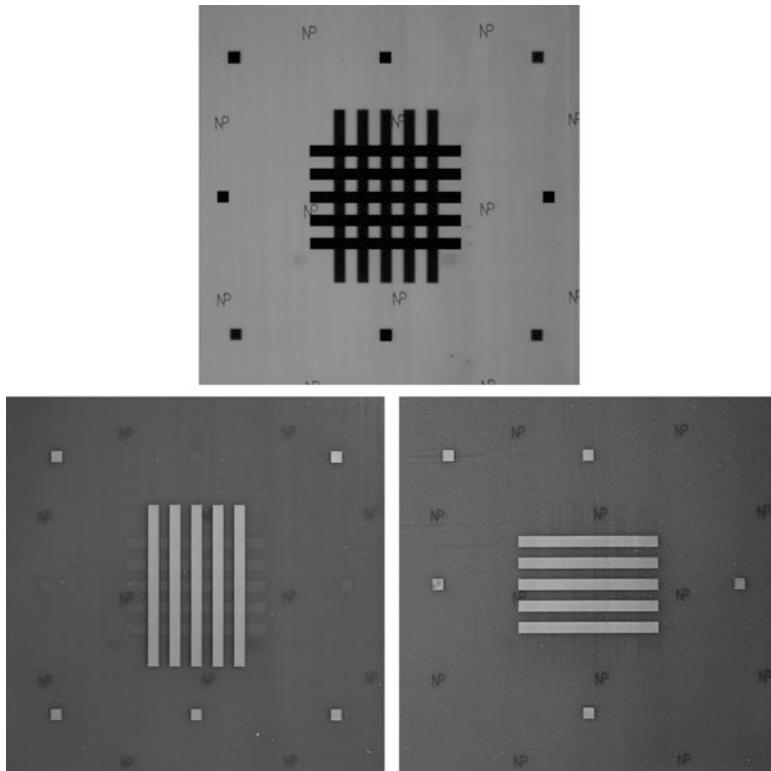
**Fig. 5** The specimen under test (left) and the reconstruction through the Monotonicity Principle. Each pixel is  $5 \times 5$  mm. (Reprinted from Tamburrino et al. 2012 with permission of Elsevier Ltd.)

Monotonicity algorithm processes the experimental data providing the reconstruction in real time during measurement data acquisition (about 100  $\mu$ s per pixel).

### Single-Face PCB

The first test is a printed circuit board where the copper (thickness 35  $\mu$ m) forms the letters IC, as shown in Fig. 5 (left).

The inversion, which is shown in Fig. 5 (right) is error-free. This is in line with the results of the numerical simulations of the whole measurement and inversion processes where we found that a target such as that of Fig. 5 can be reconstructed without errors, as long as the noise level affecting the data is smaller than a threshold of 20 m $\Omega$ . With our experimental setup, we found a noise level in the order of 10 m $\Omega$ .



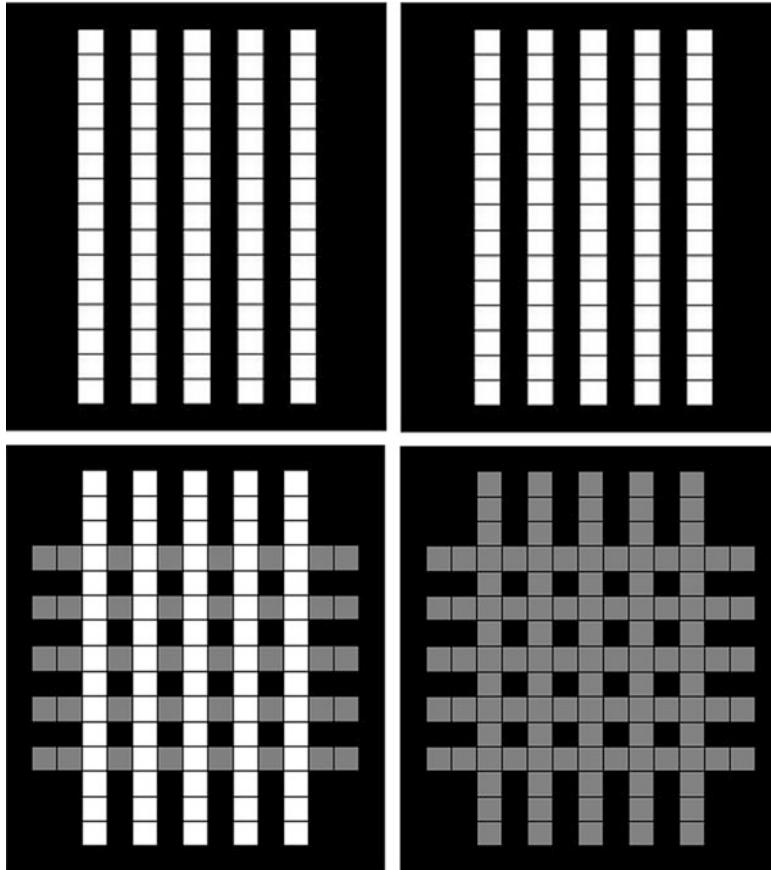
**Fig. 6** The specimen under test. Top: transparency photograph of the PCB, the copper islands on the top and bottom sides are clearly visible. Bottom: photograph of the top side (right) and photograph of the bottom side (left). The photograph of the bottom side has been flipped horizontally. (Reprinted from Tamburrino et al. 2012 with permission of Elsevier Ltd.)

### Double-Face PCB

The second test refers to the imaging of a double-sided printed circuit board. It consists of a double-sided PCB (copper, thickness 35 µm) as shown in Fig. 6.

The measurements are taken from the top side of the PCB only. In this case the imaging algorithm requires test anomalies on both sides (top and bottom) of the PCB.

This case is more intriguing than the previous one. Indeed, one expects the test anomaly for a specific side of the PCB (either top or bottom) to provide the reconstruction for the same side. This is the case for the test anomaly associated to the top side as shown in Fig. 7 (top) which presents an error-free reconstruction. Concerning the test anomaly for the bottom layer, we found that it yields a reconstruction of the bottom layer which superposes the pixels of the two sides (Fig. 7, bottom).



**Fig. 7** Top: results of processing the data from the test anomaly in the top layer. The true map of the top layer (left) and the reconstructed map (right). Bottom: results of processing the data from the test anomaly in the bottom layer. The superpositions of the pixels from the bottom (gray) and top (white) layers (left) and the reconstruction of the bottom layer (right). The test anomaly in the bottom layer yields a reconstruction of the bottom layer which superposes the two sides of the PCB. (Reprinted from Tamburrino et al. 2012 with permission of Elsevier Ltd.)

This (apparently) unexpected result can be easily explained by considering that:

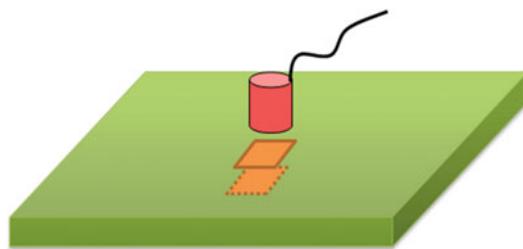
$$\mathbf{R}_{k_{\text{TOP}}} \geq \mathbf{R}_{k_{\text{BOTTOM}}} \quad (13)$$

where  $k_{\text{TOP}}$  and  $k_{\text{BOTTOM}}$  refer to two test anomalies occupying the same in-plane position but located in different sides (see Fig. 8).

Thanks to the relationship (13), it is trivial to prove that the test domains related to the bottom side provide a reconstruction that is the union of the pixels from the top and the bottom sides (Fig. 7).



**Fig. 8** The two test anomalies occupying the same in-plane position but located on different sides. (Reprinted from Tamburrino et al. 2012 with permission of Elsevier Ltd.)



**Fig. 9** The double-pixel test anomaly made by two pixels in the same in-plane position but on opposite sides. The Monotonicity test performed with this test anomaly provides the common pixels. (Reprinted from Tamburrino et al. 2012 with permission of Elsevier Ltd.)

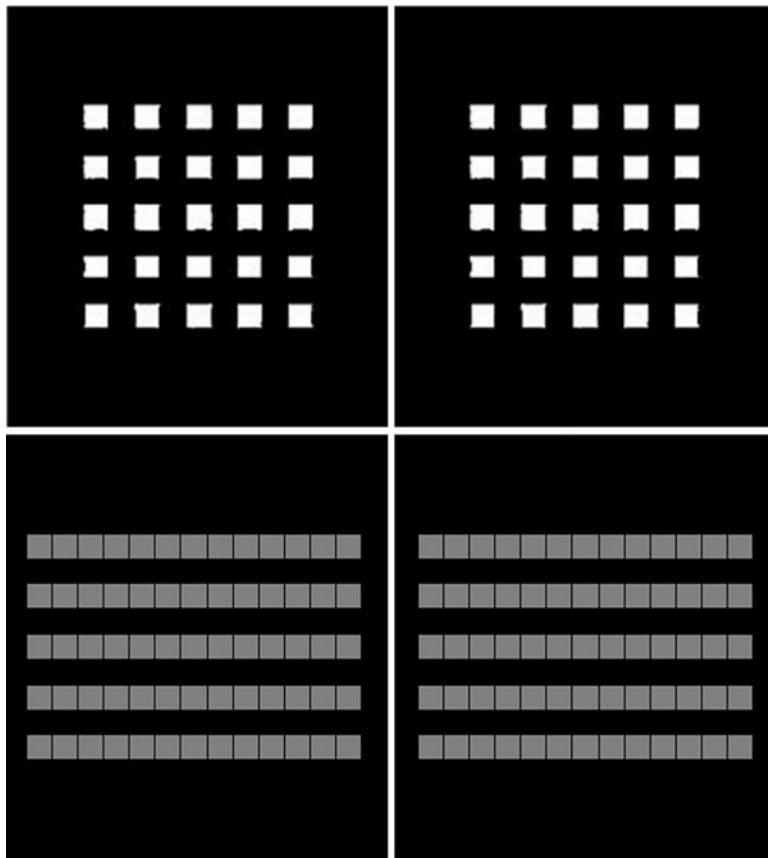
In order to properly reconstruct the bottom side, we have to identify the pixels that are common to the top and bottom layers. For doing this, we perform a third test where we introduce a test anomaly (double pixel anomaly, see Fig. 9) which presents two pixels in the same in-plane position but on opposite sides. This configuration is capable of detecting the pixel shared by the top and bottom sides as shown in Fig. 10, and, therefore, by combining these three reconstructions, it is possible to obtain an error-free reconstruction of the bottom side, as well (see Fig. 10). Additional examples of reconstruction from experimental data can be found in Tamburrino et al. (2012).

---

## Monotonicity Principle for Eddy Currents in Other Settings

### Small Skin-Depth Regime

Small skin-depth regime in the framework of eddy current imaging refers to the case when (i) the displacement current is still negligible and (ii) the skin-depth is smaller than the relevant geometrical size of the problem such as the thickness of the specimen, etc.



**Fig. 10** Top: results of processing the data from the test anomaly made by the double pixel anomaly (two pixels on opposite sides of the PCB). The true map of the common pixels for the two sides (left) and its reconstruction (right). Bottom: final results for the bottom layer obtained by combining the reconstructions from the three test anomalies. The bottom layer and its reconstruction (left). (Reprinted from Tamburrino et al. 2012 with permission of Elsevier Ltd.)

Small skin-depth regime is attractive because the signals are strong enough. This regime is appropriate for imaging the surface of a conductive material, rather than its interior (the pronounced skin effect prevents the field to penetrate inside the specimen). This is the case, for instance, when the interest is in detecting and imaging surface breaking defects or anomalies.

In this regime, the inductance matrix, that is, the imaginary part of the impedance matrix divided by  $\omega$  ( $\mathbf{L} = \text{Im}\{\mathbf{Z}\}/\omega$ ), is monotonic. Specifically, the inductance matrix is monotonic with respect to the domain occupied by the conductive specimen, because the fields are not capable to penetrate the material:

$$D_1 \subseteq D_2 \Rightarrow \mathbf{L}_{D_1} \geq \mathbf{L}_{D_2}. \quad (14)$$

In (14)  $\mathbf{L}_{D_1}$  and  $\mathbf{L}_{D_2}$  are the inductance matrices related to a conductive domain occupying region  $D_1$  and  $D_2$ , respectively (see Tamburrino et al. 2010). From the physical standpoint, (14) means that (for prescribed currents applied to the coils) the magnetic energy is higher for the configuration where the specimen occupies the smaller domain ( $D_1$ ). Monotonicity (14) can be turned into a Monotonicity for (surface breaking) defects. Indeed, from (14) we have

$$\Omega_k \subseteq V \subseteq D \Rightarrow \mathbf{L}_{D \setminus V} \geq \mathbf{L}_{D \setminus \Omega_k} \quad (15)$$

where  $D$  is the conducting domain (without defect),  $\Omega_k$  is, as usual, a test domain, and  $V$  is the unknown defect. Similarly,

$$V \subseteq \Omega_k \subseteq D \Rightarrow \mathbf{L}_{D \setminus \Omega_k} \geq \mathbf{L}_{D \setminus V}. \quad (16)$$

Propositions (15) and (16) are equivalent to (6) and (7), thus allowing to apply without any major modification the imaging method of section “[Imaging Method](#).”

### Example

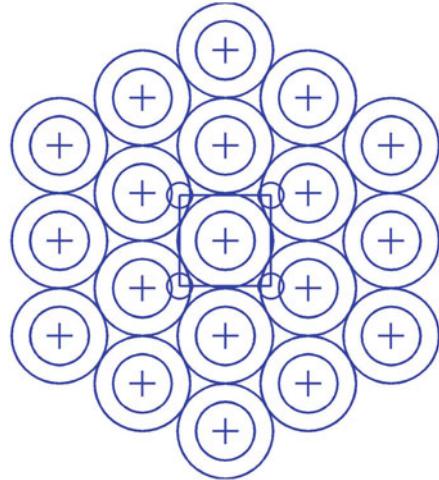
The numerical example of this section refers to the inspection of a plate (copper, electrical conductivity of  $5.88 \times 10^7$  S/m) that may present (surface breaking) volumetric defects (see Tamburrino et al. 2010 for details).

The plate is  $309 \times 309 \times 0.5$  mm. For the sake of simplicity, we assume that the defects have a constant height of 100% of the thickness of the plate, i.e., 0.5 mm, and are perfectly insulating. The plate is subdivided in a regular Cartesian grid of  $100 \times 100 \times 1$  test anomalies (each test anomaly is a perfectly insulating voxel of  $3.09 \times 3.09 \times 0.5$  mm). Defects with different heights can be easily treated by considering set of test anomalies with different heights. The skin-depth can be considered negligible for frequencies greater than 172 kHz (the skin-depth in copper at 172 kHz is 0.05 mm).

The probe is made up of a regular two-dimensional array of coils, and it can be translated to inspect large areas of the specimen. The coils are arranged in the closest packed array configuration to maximize the number of coils per area as shown in Fig. 11. As a realistic example, we consider a coil used for a set of benchmarks developed by the Laboratoire d’Intégration des Systèmes et des Technologies (LIST) at the Commissariat à l’Énergie Atomique (CEA, Saclay, France), concerning the inspection of heat exchanger tubes in the nuclear industry. The main parameters are internal diameter 2 mm, outer diameter 3.25 mm, height 2 mm, lift-off 0.3 mm, and number of turns 238. The total number of coil is 19.

The impedance matrix is collected onto a regular Cartesian grid having a step of one half of the voxel size (1.545 mm). Hereafter, we refer to  $\mathbf{L}_{D \setminus V}^m$  and  $\mathbf{L}_{D \setminus \Omega_k}^m$  as the inductance matrices for the unknown defect  $V$  and the test anomaly  $\Omega_k$  collected when the probe is at position  $m$ . Also, we define  $\delta\mathbf{L}_{D \setminus V}^m = \mathbf{L}_{D \setminus V}^m - \mathbf{L}_D^m$  and  $\delta\mathbf{L}_{D \setminus \Omega_k}^m = \mathbf{L}_{D \setminus \Omega_k}^m - \mathbf{L}_D^m$ , being  $\mathbf{L}_D^m$  the data corresponding to the defect-free configuration.  $\delta\mathbf{L}_{D \setminus V}^m$  and  $\delta\mathbf{L}_{D \setminus \Omega_k}^m$  represent the “useful” part of the measurement or data for the test domain, respectively. The noise is additive and its level is 11% of the mean of the

**Fig. 11** The probe made by an array of 19 identical coils. The external radius of the probe is equal to the (in-plane) size of a voxel. (Reprinted from Tamburrino et al. 2010 with the permission of IOP Publishing Ltd.)



norm of the  $\delta\tilde{\mathbf{L}}_{D \setminus V}^m$ 's. The (noisy) measured data is  $\delta\tilde{\mathbf{L}}_{D \setminus V}^m = \delta\tilde{\mathbf{L}}_{D \setminus V}^m + \mathbf{N}^m$ , being  $\mathbf{N}^m$  the noise matrix. Measurements from multiple positions contain more information than from a single position. Combining properly these multiple measurements is nontrivial. In Tamburrino et al. (2010), we discussed different strategies for combining multiple measurements. The more effective consists in defining the sign index for a specific voxel as the minimum among those obtained from the positions where the probe interact with the prescribed voxel:

$$S_k^{\min} = \min_{m \in E_k} S_k^m$$

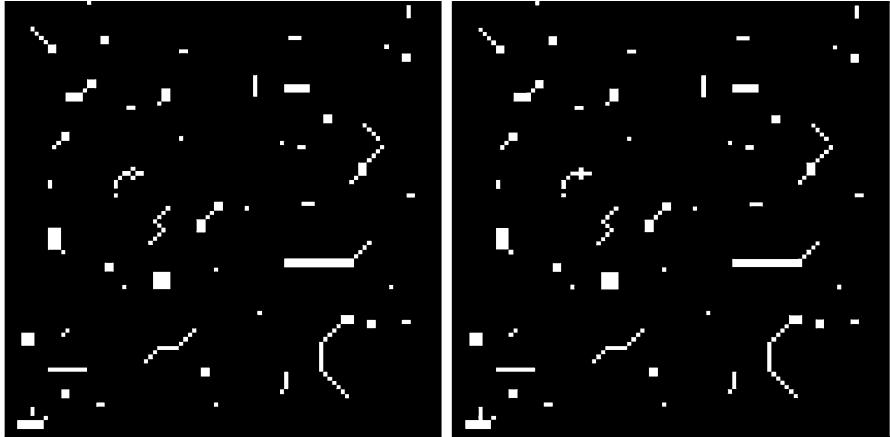
where  $E_k = \{m \mid \text{the voxel } k \text{ interacts with the probe in position } m\}$ . Moreover, we found an enhanced sensitivity when the sign index

$$S_k^m = \frac{\sum_j \lambda_j (\delta\tilde{\mathbf{L}}_{D \setminus V}^m - \delta\mathbf{L}_{D \setminus \Omega_k}^m)}{\sum_j |\lambda_j (\delta\tilde{\mathbf{L}}_{D \setminus V}^m - \delta\mathbf{L}_{D \setminus \Omega_k}^m)|}$$

is evaluated by discarding those eigenvalues which are smaller than the norm of the noise matrix  $\|\mathbf{N}^m\|$ . The reconstruction related to the example of (Tamburrino et al. 2010) is shown in Fig. 12.

### Pulsed Eddy Current Imaging: Time Constants

In a pulsed eddy current imaging method, the source is switched off after an initial pulse. The pulse usually presents sharp edges (small rise and fall time), in order to be rich in terms of frequency content. The response of the system is measured in the time domain. After the initial pulse is over, the response can be easily written in terms of time constants  $\tau_1, \tau_2, \dots$ . For instance, the induced eddy current density can be represented as



**Fig. 12** True map (left) and its reconstruction (right). (Reprinted from Tamburrino et al. 2010 with the permission of IOP Publishing Ltd.)

$$\mathbf{J}(\mathbf{r}, t) = \sum_{i=1}^{+\infty} c_i \mathbf{J}_i(\mathbf{r}) e^{-t/\tau_i}, \quad (17)$$

the corresponding magnetic flux density as

$$\mathbf{B}^{\text{eddy}}(\mathbf{r}, t) = \sum_{i=1}^{+\infty} c_i \mathbf{B}_i^{\text{eddy}}(\mathbf{r}) e^{-t/\tau_i},$$

and the induced voltage across the  $k$ -th coil as

$$v_k^{\text{eddy}}(t) = \frac{d\phi_k^{\text{eddy}}(t)}{dt} = \sum_{i=1}^{+\infty} c_i v_{k,i} e^{-t/\tau_i}.$$

We notice that  $\mathbf{J}(\mathbf{r}, t)$ ,  $\mathbf{B}^{\text{eddy}}(\mathbf{r}, t)$ , and  $v^{\text{eddy}}(t)$  share the same time constants. In this way, it is possible using existing pulsed eddy current testing system without the need of modifying them. Moreover  $\mathbf{B}^{\text{eddy}}(\mathbf{r}, t)$  or  $v_k^{\text{eddy}}(t)$ , which be experimentally measured in a nondestructive manner, provide the waveforms from which one has to estimate the time constants for the Monotonicity-based imaging method.

The set of time constants  $\{\tau_1, \tau_2, \dots\}$ , when ordered, is monotonic with respect to the electrical conductivity of the specimen under testing. Assuming the time constants ordered in decreasing order, it has been proved that (Su et al. 2017b)

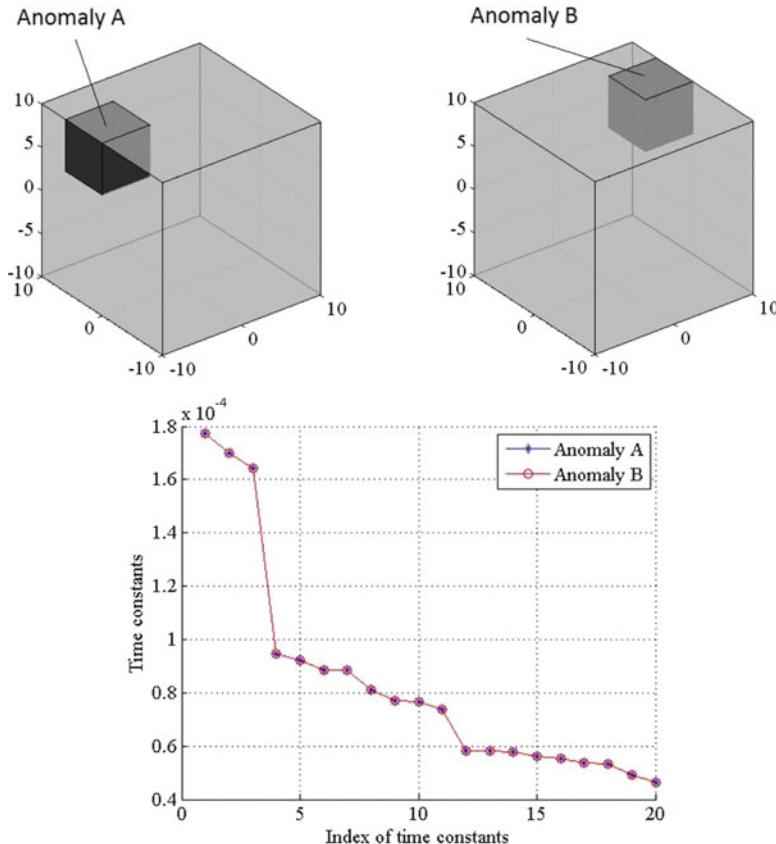
$$\eta_1(\mathbf{r}) \leq \eta_2(\mathbf{r}) \quad \forall \mathbf{r} \in D \Rightarrow \tau_{\eta_1}^i \geq \tau_{\eta_2}^i \quad \forall i \in \mathbb{N}. \quad (19)$$

This Monotonicity Proposition is similar to (2), and, therefore, we can apply again MPM. From (19) it follows the equivalent of (4);

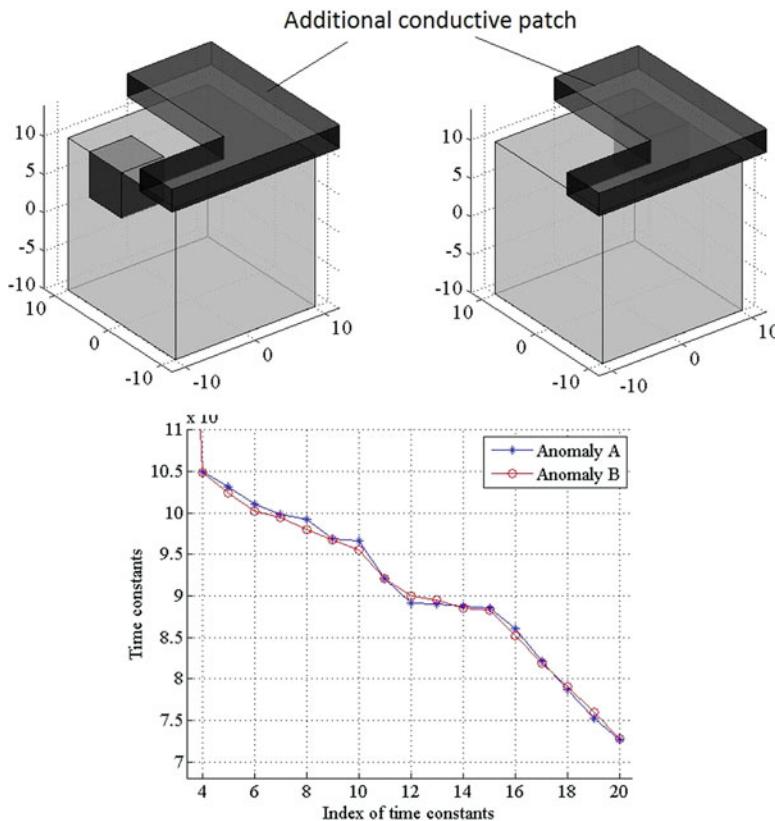
$$V_1 \subseteq V_2 \subseteq D \Rightarrow \tau_{V_1}^i \geq \tau_{V_2}^i \quad \forall i \in \mathbb{N},$$

which is useful for imaging defect in an otherwise homogeneous background.

Monotonicity Principle in (19) is somehow different from previous MP. Specifically, it involves quantities (time constants) related to the free modes of the system. These modes do not depend on the probing system, and, therefore, they are immune to lift-off and tilting errors. This is a key advantage because, as it is well known, lift-off and tilting may introduce significant errors. However, this introduces a drawback: two configurations, which differ for an isometry of the space only, cannot be distinguished from the knowledge of the time constants (see Fig. 13). However, by introducing an additional conductive patch (Su et al. 2017b), this kind of symmetry is broken, and the two anomalies can be distinguished from the knowledge of the time constants (see Fig. 14).



**Fig. 13** A  $20 \times 20 \times 20$  mm cube of resistivity of  $10^{-7} \Omega\text{m}$  with two possible anomalies A and B (Top). The two configurations differ by an isometry only. The related time constants (Bottom). (Reprinted from Su et al. 2017b with the permission of IOP Publishing Ltd.)



**Fig. 14** The same  $20 \times 20 \times 20$  mm cube of resistivity of  $10^{-7} \Omega\text{m}$  with two possible anomalies A and B plus the additional patch (Top). The related time constants (Bottom). (Reprinted from Su et al. 2017b with the permission of IOP Publishing Ltd.)

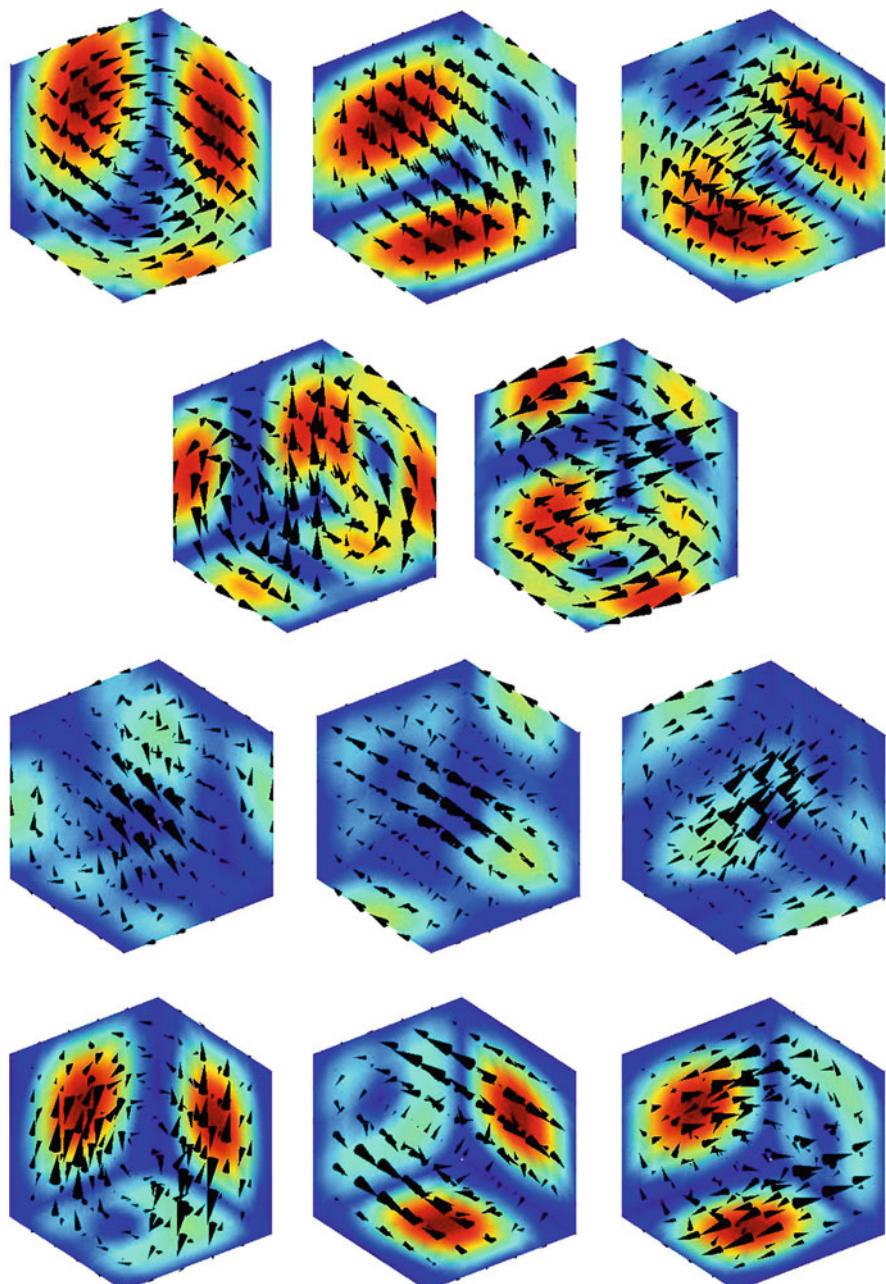
Time constants are related to natural modes. In terms of eddy current density, the  $i$ -th time constant is related to the natural mode  $\mathbf{J}_i(\mathbf{r})$  (see 17).

The natural modes are global, i.e., distributed inside the conductive domain as a whole (Su et al. 2017b); see Fig. 15.

This is advantageous because it makes possible to inspect, to some extent, also the innermost region of the conductive materials. On the other hand, modes are global, and, therefore, a defect needs to be of the proper size to interact with a mode.

## Summary

This chapter discusses the main features of the Monotonicity Principle and the related imaging method (MPM) in eddy current tomography. MPM is a noniterative imaging method where a voxel can be evaluated independently from all the others,



**Fig. 15** The first 11 modes for a cube. The modes number is in increasing order from top to bottom and from left to right. (Reprinted from Su et al. 2017b with the permission of IOP Publishing Ltd.)

thanks to a proper local (voxel-by-voxel) and computational cheap test. This makes the method very fast, efficient, and suitable for real-time imaging.

MPM possess some unique features: (i) the local test is not based on the minimization of a proper cost function and, therefore, MPM does not suffer from being trapped in local minima; (ii) it is capable to provide upper and lower bounds to the unknown quantity, even in the presence of noise; and (iii) MP provides a rather general framework which can be customized to specific needs such as imaging, sizing, or material characterization.

MP can been applied to ECT in different settings, from frequency domain (either small or large skin-depth regimes) to time-domain. Moreover, MP constitutes a rather general approach for facing inverse problems from different areas: other than inverse problems arising from parabolic PDEs, such as eddy current tomography, MP can also be applied to inverse problems governed by elliptic and hyperbolic PDEs.

**Acknowledgments** Part of this work has been reproduced from (Tamburino et al., 2010), (Tamburino et al., 2012), (Su et al., 2017b) and (Tamburino et al., 2016) with the permission of the Publishers. The Authors are grateful to the Publishers.

---

## References

- Abu-Nabah BA, Nagy PB (2006) Iterative inversion method for eddy current profiling of near-surface residual stress in surface-treated metals. *NDT & E Int* 39(8):641–651
- Ammari H, Iakovleva E, Lesselier D (2005) A music algorithm for locating small inclusions buried in a half-space from the scattering amplitude at a fixed frequency. *Multiscale Model Simul* 3(3): 597–628
- Aykroyd RG, Soleimani M, Lionheart WRB (2005) Full shape reconstruction from partial ERT monotonicity information using a Bayes-MCMC approach. In: Proceedings of the 4th world congress in industrial process tomography, International Society for Industrial Process Tomography, Japan, pp 697–702
- Badics Z, Pávó J, Komatsu H, Kojima S, Matsumoto Y, Aoki K (1998) Fast flaw reconstruction from 3D eddy current data. *IEEE Trans Magn* 34(5):2823–2828
- Bowler J (2002) Thin-skin eddy-current inversion for the determination of crack shapes. *Inverse Probl* 18(6):1891–1905
- Calvano F, Caturano G, Cavaccini G, Rubinacci G, Tamburino A, Ventre S (2010) Crack shape reconstruction from ect using topology constrained optimization (TOPCSA) algorithms. *Stud Appl Electromagn Mech* 33:230–237
- Colton D, Kirsch A (1996) A simple method for solving inverse scattering problems in the resonance region. *Inverse Probl* 12(4):383–393
- Devaney AJ (2000) Super-resolution processing of multi-static data using time reversal and music. Northeastern University Preprint. [http://ece.neu.edu/fac-ece/devaney/preprints/paper02n\\_00.pdf](http://ece.neu.edu/fac-ece/devaney/preprints/paper02n_00.pdf)
- Dorn O, Lesselier D (2009) Level set methods for inverse scattering – some recent developments. *Inverse Probl* 25(12):125001
- Flores-Tapia D, Pistorius S (2010) Electrical impedance tomography reconstruction using a monotonicity approach based on a priori knowledge. In: 2010 Annual international conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10, pp 4996–4999
- Flores-Tapia D, O'Halloran M, Pistorius S (2011) A bimodal reconstruction method for breast cancer imaging. *Prog Electromagn Res* 118:461–486
- Garde H (2018) Comparison of linear and non-linear monotonicity-based shape reconstruction using exact matrix characterizations. *Inverse Probl Sci Eng* 26(1):33–50

- Garde H, Staboulis S (2017) Convergence and regularization for monotonicity-based shape reconstruction in electrical impedance tomography. *Numer Math* 135(4):1221–1251
- Hanke M, Brühl M (2003) Recent progress in electrical impedance tomography. *Inverse Probl* 19(6):S65–S90
- Harrach B, Ullrich M (2013) Monotonicity-based shape reconstruction in electrical impedance tomography. *SIAM J Math Anal* 45(6):3382–3403
- Harrach B, Ullrich M (2015) Resolution guarantees in electrical impedance tomography. *IEEE Trans Med Imaging* 34(7):1513–1521
- Harrach B, Lee E, Ullrich M (2015) Combining frequency-difference and ultrasound modulated electrical impedance tomography. *Inverse Probl* 31(9):095003
- Isakov V (1993) Uniqueness and stability in multi-dimensional inverse problems. *Inverse Probl* 9(6):579–621
- Kirsch A (1998) Characterization of the shape of a scattering obstacle using the spectral data of the far field operator. *Inverse Probl* 14(6):1489–1512
- Li Y, Udfa L, Udfa SS (2004) Three-dimensional defect reconstruction from eddy-current NDE signals using a genetic local search algorithm. *IEEE Trans Magn* 40(2):410–417
- Lionheart WRB, Soleimani M, Peyton AJ (2003) Sensitivity analysis of 3D magnetic induction tomography (MIT). In: Proceedings of the 3rd world congress on industrial process tomography, International Society for Industrial Process Tomography, Canada, pp 239–244
- Luong B, Santosa F (1998) Quantitative imaging of corrosion in plates by eddy current methods. *SIAM J Appl Math* 58(5):1509–1531
- Maffucci A, Vento A, Ventre S, Tamburrino A (2016) A novel technique for evaluating the effective permittivity of inhomogeneous interconnects based on the monotonicity property. *IEEE Transactions on Components, Packaging and Manufacturing Technology* 6(9):1417–1427
- Nair S, Rose JH (1990) Reconstruction of three-dimensional conductivity variations from eddy current (electromagnetic induction) data. *Inverse Probl* 6(6):1007
- Norton SJ, Bowler JR (1993) Theory of eddy current inversion. *J Appl Phys* 73(2):501–512
- Pirani A, Ricci M, Specogna R, Tamburrino A, Trevisan F (2008) Multi-frequency identification of defects in conducting media. *Inverse Probl* 24(3):035011
- Premel D, Baussard A (2002) Eddy-current evaluation of three-dimensional flaws in flat conductive materials using a bayesian approach. *Inverse Probl* 18(6):1873
- Qi-Nian J (2000) On the iteratively regularized gauss-newton method for solving nonlinear ill-posed problems. *Math Comput* 69(232):1603–1623
- Soleimani M, Lionheart WRB (2006) Absolute conductivity reconstruction in magnetic induction tomography using a nonlinear method. *IEEE Trans Med Imaging* 25(12):1521–1530
- Stolzenburg JC, Doane JW, Jarzynski J, Jacobs LJ (2003) Near field inversion method to measure the material properties of a layer. *NDT & E Int* 36(7):523–533
- Su Z, Udfa L, Giovinco G, Ventre S, Tamburrino A (2017a) Monotonicity principle in pulsed eddy current testing and its application to defect sizing. In: Proceedings of the 2017 International Applied Computational Electromagnetics Society (ACES) symposium, Institute of Electrical and Electronics Engineers Inc., Italy, pp 1–2
- Su Z, Ventre S, Udfa L, Tamburrino A (2017b) Monotonicity based imaging method for time-domain eddy current problems. *Inverse Probl* 33(12):125007
- Su Z, Ye C, Giovinco G, Ventre S, Udfa L, Tamburrino A (2018) Evaluation of electrical conductivity of metals via monotonicity of time constants. *Stud Appl Electromagn Mech* 43:88–95
- Tamburrino A (2006) Monotonicity based imaging methods for elliptic and parabolic inverse problems. *J Inverse Ill-Posed Probl* 14(6):633–642
- Tamburrino A, Rubinacci G (2002) A new non-iterative inversion method for electrical resistance tomography. *Inverse Probl* 18(6):1809–1829
- Tamburrino A, Rubinacci G (2006) Fast methods for quantitative eddy-current tomography of conductive materials. *IEEE Trans Magn* 42(8):2017–2028

- Tamburri A, Ventre S, Rubinacci G (2010) Recent developments of a monotonicity imaging method for magnetic induction tomography in the small skin-depth regime. *Inverse Probl* 26(7):074016
- Tamburri A, Calvano F, Ventre S, Rubinacci G (2012) Non-iterative imaging method for experimental data inversion in eddy current tomography. *NDT & E Int* 47:26–34
- Tamburri A, Barbato L, Colton D, Monk P (2015) Imaging of dielectric objects via monotonicity of the transmission eigenvalues. In: Proceedings of the 12th international conference on mathematical and numerical aspects of wave propagation, Karlsruhe, pp 99–100
- Tamburri A, Vento A, Ventre S, Maffucci A (2016) Monotonicity imaging method for flaw detection in aeronautical applications. *Stud Appl Electromagn Mech* 41:284–292
- Wallinger C, Watzenig D, Steiner G, Brandstätter B (2009) Adaptive monotonicity method for permittivity imaging. *COMPEL* 28(4):892–906
- Yamamoto M (1997) A mathematical aspect of inverse problems for non-stationary Maxwell's equations. *Int J Appl Electromagn Mech* 8(1):77–98
- Yin W, Dickinson SJ, Peyton AJ (2005) Imaging the continuous conductivity profile within layered metal structures using inductance spectroscopy. *IEEE Sensors J* 5(2):161–166
- Yusa N, Cheng W, Uchimoto T, Miya K (2002) Profile reconstruction of simulated natural cracks from eddy current signals. *NDT & E Int* 35(1):9–18
- Yusa N, Chen Z, Miya K, Uchimoto T, Takagi T (2003) Large-scale parallel computation for the reconstruction of natural stress corrosion cracks from eddy current testing signals. *NDT & E Int* 36(7):449–459



# Motion-Induced Eddy Current Testing

22

Hartmut Brauer and Marek Ziolkowski

## Contents

Introduction .....	782
Lorentz Force Eddy Current Testing (LET) .....	784
Theory .....	786
Comparison of ECT and LET .....	788
Motion-Induced Eddy Current Testing (MIECT) .....	790
Numerical Simulations .....	790
Computation of Eddy Current Distributions in Moving Parts .....	791
Numerical Modeling .....	792
Comparison of Numerical Approaches .....	798
Measurement and Sensor System .....	799
Force Measurement .....	799
Differential Lorentz Force Eddy Current Testing Sensor .....	800
Optimization of the Permanent Magnet System .....	801
Lorentz Force Evaluation .....	803
Applications .....	806
Inspection of Glass Laminate Aluminum Reinforced Epoxy .....	807
Defectoscopy of Friction Stir Welding .....	811
Potential Applications of MIECT .....	819
References .....	821

---

H. Brauer (✉)

Technische Universitaet Ilmenau, Ilmenau, Germany

e-mail: [hartmut.brauer@tu-ilmenau.de](mailto:hartmut.brauer@tu-ilmenau.de)

M. Ziolkowski

Technische Universitaet Ilmenau, Ilmenau, Germany

West Pomeranian University of Technology Szczecin, Szczecin, Poland

e-mail: [marek.ziolkowski@tu-ilmenau.de](mailto:marek.ziolkowski@tu-ilmenau.de); [marek.ziolkowski@zut.edu.pl](mailto:marek.ziolkowski@zut.edu.pl)

**Abstract**

Nondestructive material testing and evaluation is a vast interdisciplinary field as well as a challenge due to the variety of applications. Whereas the focus of nondestructive testing is to identify anomalies within a specimen, the reconstruction of defect properties and their influence on the materials usability is the focus of nondestructive evaluation. In this chapter the technology of motion-induced eddy current testing (MIECT) is introduced. In contrast to traditional eddy current testing (ECT) methods, MIECT makes use of relative motion between the object under test and permanent magnets. The induced eddy currents interact with the applied magnetic field and result in a Lorentz force, depending on the impressed magnetic induction, the electrical conductivity, and the measuring velocity. Because permanent magnets produce considerably stronger magnetic fields than current-carrying ECT coils, even deep internal defects can be detected using the Lorentz force eddy current testing (LET). It is shown how the electromagnetic fields can be described theoretically and simulated numerically, as well as how imperfections/defects in non-ferromagnetic, conducting specimens can be detected using an appropriate laboratory environment. Comparative studies have shown that LET applied to metallic composite material or friction stir welds is a promising and competitive alternative to traditional ECT methods enabling the contactless evaluation of moving electrical conductors.

---

**Introduction**

Nondestructive testing (NDT) of materials and products is of great interest in a variety of modern engineering applications. NDT in general enables the initial inspection of test samples to confirm the structural integrity of safety-relevant components without causing damage.

This chapter focuses on electromagnetic NDT methods and more specifically on the motion-induced eddy current testing (MIECT). Traditional eddy current testing methods (ECT) make use of time-dependent magnetic fields to induce eddy currents in the object under test. Those are altered in the presence of physical irregularities such as flaws, cracks, or inclusions. In ECT, the variations are detected by measuring the magnetic flux through a pickup coil produced by the disturbed eddy currents in the specimen. In contrast to the ECT methods, MIECT makes use of a relative motion between the object under test and a permanent magnet configuration. The induced eddy currents interact with the applied magnetic field and result in a Lorentz force. Considering Newton's third law, the force acts on both, the specimen and the magnet itself, where it is measured. This quantity is used to evaluate the integrity of the structure under test. Thus, MIECT differs from traditional ECT methods in the way how eddy currents are induced and how signals are evaluated. It is the intent of this chapter to introduce the technology of motion-induced eddy current testing.

The Lorentz force eddy current testing (LET) method, a rather new MIECT method, was invented to complement established methods by overcoming the well-known detection limitations for subsurface defects. Originally, the MIECT

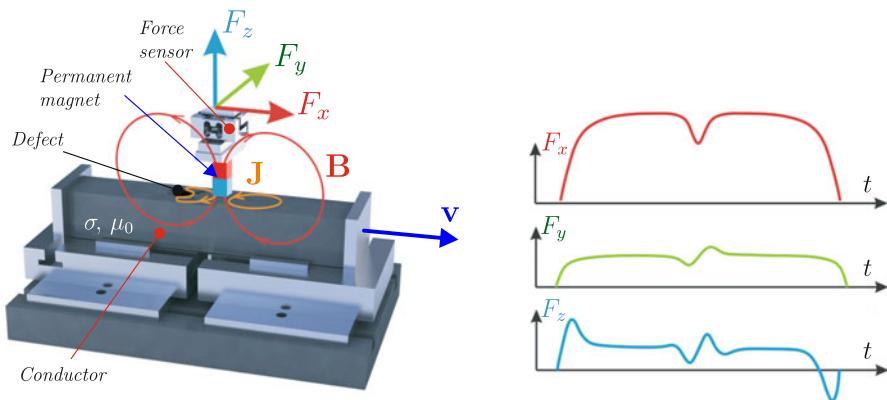
method was proposed in (Brauer and Ziolkowski 2008). Important advantages lie in the application of magnetostatic fields, which potentially allow the detection of defects lying deep inside the object under test. The working principle of LET permits the inspection of moving parts as can frequently be found in industrial settings. This necessitates appropriate NDT methods like LET, which are capable of testing moving objects. In the past more than 10 years, the authors have been conducting basic research by means of considerable funding from the German Research Foundation. This support ensures the continuous improvement of the related technologies in theory and academic practice but, so far, only allowing performance of basic research and experimental studies in university labs.

In recent years, an increase of a variety of methods can be observed which make use of relative motion between a magnetic field source producing a stationary magnetic field and the object under test. Techniques based on this principle can be classified as motion-induced eddy current testing (MIECT) methods.

The group of Chady et al. has recently realized a prototype equipment consisting of an eddy current transducer and rotating permanent magnets (Chady and Spychaliski 2017). The eddy current transducer is designated for testing of planar conducting plates. A rotating head with permanent magnets is used to induce eddy currents in the specimen. Two Hall-effect devices connected in a differential manner are utilized to measure the eddy current reaction. This inspection system is effective especially in cases of thick metallic elements, when it is necessary to utilize low excitation frequency, or in systems without power supply. Thus, the system is suitable for applications where it is important to achieve high penetration depth.

The group of Ribeiro and Ramos proposed in 2013 with the velocity-induced eddy current testing (VIECT) another approach where the magnetic fields are also measured (Geirinhas Ramos et al. 2013). Ramos et al. investigated the applicability of moving stationary magnetic field sources using DC coils in the framework of NDT (Ramos et al. 2013; Rocha 2017). They measured the disturbances of the magnetic field resulting from a defect directly by means of giant magneto-resistive (GMR) sensors.

Moreover, they also exchanged the magnetic field source with a permanent magnet in order to achieve higher flux densities and an increased induced eddy current density inside the moving specimen. The sensor orientation has to be chosen carefully when using GMR sensors in order to avoid saturation effects. This can be overcome by applying differential coils as magnetic field sensors as it is also done in the framework of ECT. In this way, only the temporal change of the magnetic flux resulting from a passing defect is measured. Rocha et al. extended the analysis to the application of Hall sensors instead of GMRs and pickup coils (Rocha et al. 2015b). They also investigated the defect response signals for different permanent magnet configurations and proposed the use of sensor arrays to expedite the assessment of larger areas. In a subsequent study, the application of GMRs, differential coils, and Hall sensors was compared in the framework of MIECT (Rocha et al. 2015a). As a result, it turned out that GMRs were able to detect defects when crossing the edges of the defect. In contrast, pickup coils and Hall sensors also provided signals when the probe passed the defect in its centerline.



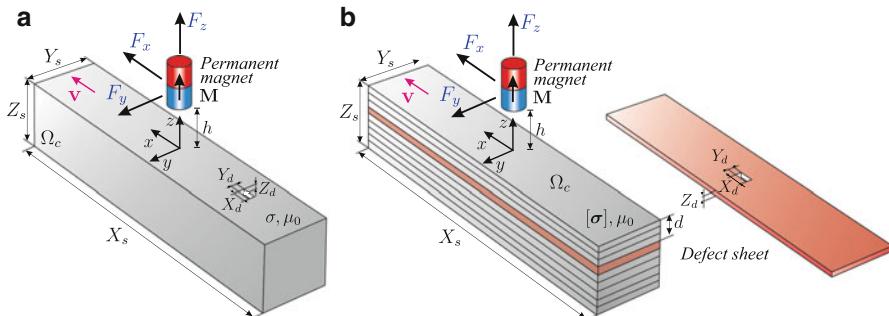
**Fig. 1** Lorentz force eddy current testing for contactless evaluation of electrically conducting materials

In 2015, another MIECT technique was proposed in (Tan et al. 2015). In contrast to previous studies which made use of translational motion, they proposed a system using rotational motion of the magnetic field source to induce eddy currents inside the object under test. A diametral magnetized, cylindrical permanent magnet rotates in close vicinity of a conductive object, and anomalies are analyzed by means of the variations in the electromagnetic torque. The use of rotational motion provides the opportunity to design portable MIECT systems in line with ECT devices currently available.

The presented studies are all limited to the analysis of surface-breaking defects indicating the early state of MIECT systems. However, it is emphasized that MIECT is not restricted to detect flaws on the surface. As in (Brauer and Ziolkowski 2008) is shown, the MIECT system consisting of a permanent magnet combined with a force sensor can be used to detect defects in conducting, nonmagnetic specimen. The principal setup is shown in Fig. 1. Because the Lorentz forces are measured, this method has been called Lorentz force eddy current testing (LET) (Brauer et al. 2014).

## Lorentz Force Eddy Current Testing (LET)

Lorentz force eddy current testing (LET) belongs to the group of MIECT-type methods. It is a technique for nondestructive and contactless evaluation of electrically conducting specimens. The basic principle, shown in Fig. 2, is based on the interaction between a permanent magnet and a moving specimen. As a consequence of this motion, eddy currents are induced inside the object under test, which in turn react with the magnetic field, producing a Lorentz force acting on both, the specimen and the permanent magnet. The novelty of the method lies in the determination of the measurement signal. In contrast to ECT and other MIECT techniques, the force



**Fig. 2** General principle of Lorentz force eddy current testing for contactless evaluation of electrically conducting material. The specimens and the geometrical parameters of the LET problem under investigation are shown in (a) for solids and in (b) for layered structures

acting on the magnet is measured using force sensors. In the presence of a defect, the eddy current profile and, hence, the resulting Lorentz force are perturbed. The physical principle of LET is an analogy to Lorentz force velocimetry (LFV) (Thess et al. 2006). In LFV, the main goal is to determine the flow rate of a conducting liquid by means of the Lorentz force which is proportional to the velocity of the liquid (Thess et al. 2007).

LET was initially demonstrated as an alternative NDT method in (Brauer and Ziolkowski 2008). The same authors in (Ziolkowski and Brauer 2010) tackled the numerical analysis of the reported experimental setup and proposed techniques to analyze the electromagnetic field problem with increased computational efficiency. The work on LET in an experimental and numerical framework was continued by Uhlig and Zec (Uhlig 2014; Zec et al. 2013).

A very important study is related to the investigation of the effect of defect depth on the Lorentz force signals exerting on the magnet. To study the impact of the defect depth, a layered specimen containing a number of aluminum sheets of the same thickness (usually 2 mm) was used. The defect depth can be changed easily if the position of the layer containing the defect is modified (Fig. 2).

A demonstrative model of LET is proposed and investigated in (Uhlig et al. 2012a). It consists of a modification to the well-known creeping magnet experiment, where a permanent magnet is slowly falling down a copper pipe (Donoso et al. 2011). The modification in this study consists of adding defects into the pipe wall such that the eddy current distribution and Lorentz force profile are disturbed. The LET method is extended to the determination of the electrical conductivity of the specimen assuming that the object under test is free of defects (Uhlig et al. 2012b). This technique is called Lorentz force sigmometry (LoFoS). It is shown that the lift-to-drag ratio of the Lorentz force components is proportional to the conductivity of the specimen such that  $\sigma = \alpha F_z / F_x$  with a calibration factor  $\alpha$  which is determined experimentally. Besides the mentioned investigations, fundamental studies exist on the influence of the Lorentz force on geometrical parameters such as the lift-off distance, the size of the magnet, as well as the size and depth of the defect (Uhlig

et al. 2011; Zec et al. 2014). These studies were accomplished by analyzing how the velocity or conductivity affects the resulting Lorentz force profile.

The state of the art of LET has been summarized in (Brauer et al. 2014). It includes a summary of the experimental setup, the numerical modeling techniques, and currently the applied defect reconstruction methods. The investigations on the forward models were supported in (Petković et al. 2013), addressing the inverse problem, i.e., the identification of the defect called Lorentz force evaluation (LFE). They proposed reconstruction algorithms to determine the shape and the location of the defects solely out of the Lorentz force profiles. There were following up several studies of the LFE problem, i.e., considering the defect identification as an extension and improvement of this first approach (Mengelkamp et al. 2015, 2016, 2017; Petković et al. 2013).

## Theory

Nondestructive testing (NDT) and nondestructive evaluation (NDE) of electrically conductive objects require reliable methods to detect material anomalies or deep-lying defects. Besides radiographic, ultrasonic, or optical techniques, electromagnetic methods such as eddy current testing (ECT) find a wide range of applications due to low-cost, easy-to-use equipment and low demands on the measurement environment (Hellier 2013; Jiles 1990). However, one of the most limiting factors in ECT is the frequency-dependent skin depth (Brauer and Ziolkowski 2008). This restricts the capability to detect deep-lying defects. With Lorentz force eddy current testing (LET), a novel electromagnetic nondestructive testing technique is presented (Thess et al. 2006; Uhlig et al. 2012a, b; Zec et al. 2013) with the aim to overcome this limitation. Lorentz force eddy current testing is based on setting an electrically conductive specimen into relative motion to a constant magnetic field. Due to Ohm's law for moving conductors, eddy currents are induced in the conductor under test

$$\mathbf{J} = \sigma \left( -\frac{\partial \mathbf{A}}{\partial t} - \nabla \varphi + \mathbf{v} \times \mathbf{B} \right) \quad (1)$$

where  $\mathbf{J}$  denotes the induced current density,  $\varphi$  the scalar electric potential,  $\mathbf{A}$  the magnetic vector potential ( $\mathbf{B} = \nabla \times \mathbf{A}, \nabla \cdot \mathbf{A} = 0$ ),  $\mathbf{v}$  the conductor velocity, and  $\mathbf{B}$  the total magnetic flux density.  $\mathbf{B}$  can be divided into a primary magnetic field (caused by a permanent magnet) and a secondary magnetic field generated by the eddy currents. The interaction of the constant magnetic field and the induced eddy currents results in a Lorentz force  $\mathbf{F}^{(L)}$  acting on the specimen. Due to Newton's third law, an equal force  $\mathbf{F}^{(PM)}$  is exerted on the permanent magnet in the opposite direction

$$\mathbf{F}^{(PM)} = -\mathbf{F}^{(L)} = \int_{V_c} \mathbf{J} \times \mathbf{B} dV \quad (2)$$

with  $V_c$  describing the volume of the specimen. If a defect is present in the conductive material, perturbations in the measured Lorentz force occur. Based on these perturbations, the defect can be detected and perhaps reconstructed.

In contrast to LET, common eddy current testing uses a time-changing current in a primary coil which generates a time-changing primary magnetic field  $\mathbf{B}^{(p)}$ . Usually, the signal used to evaluate the material is the change in impedance of the secondary coil.

Both principles are based on the induction of eddy currents, whereas major differences arise in shape and magnitude of the induced current densities as well as in the method of signal evaluation. Figure 3 shows a comparison of the two methods and illustrates the perturbation of eddy currents due to defects. In both methods, a secondary magnetic field  $\mathbf{B}^{(s)}$  is generated which interacts with the primary magnetic field  $\mathbf{B}^{(p)}$ . The total magnetic field is given by the sum of the two fields:

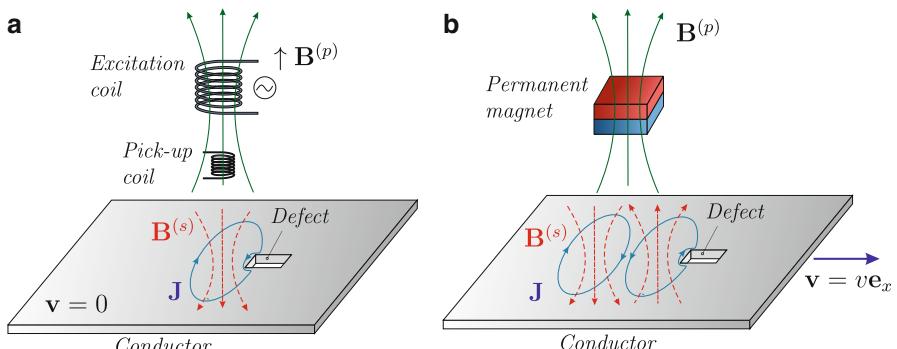
$$\mathbf{B} = \mathbf{B}^{(p)} + \mathbf{B}^{(s)}.$$

The formalism to describe the LET and ECT problem in theory is given by the magnetic convection diffusion equation (Zec 2013; Zec et al. 2013), which can be written in its potential form as

$$\nabla \times \left( \frac{1}{\mu_0} \nabla \times \mathbf{A} - \mathbf{M} \right) = -\sigma \left( \frac{\partial \mathbf{A}}{\partial t} + \nabla \varphi - \mathbf{v} \times \nabla \times \mathbf{A} \right) + \mathbf{J}^{(e)}, \quad (3)$$

if a linear and non-ferromagnetic material was assumed. In (3),  $\mathbf{M}$  denotes the magnetization vector,  $\mathbf{J}^{(e)}$  the external current density, and  $\mathbf{v}$  the velocity of the object under test. The limiting factor of ECT is the skin depth  $\delta = \sqrt{2/\omega\sigma\mu_0}$ , which results in a fast decay of the information signal for subsurface defects.

A similar factor, namely, the magnetic Reynolds number  $R_m$ , can be defined for moving conductors. By transforming the magnetic convection diffusion equation into its nondimensional form, the Reynolds number can be derived (Uhlig et al. 2012a):



**Fig. 3** Comparison of characteristic eddy current profiles observed in ECT and LET: (a) in ECT, the eddy currents show a circular profile, which are similar to an imprint of the primary excitation coil; (b) in LET, the eddy currents follow a figure of eight and do cross below the magnet

$$R_m = \mu\sigma | \mathbf{v} | L. \quad (4)$$

The parameter  $L$  is the typical length scale of the problem. In general, for  $R_m \ll 1$  diffusion of the magnetic field dominates, and the resulting field is primarily determined by the boundary conditions and the primary magnetic field  $\mathbf{B}^{(p)}$ . For  $R_m \gg 1$ , the magnetic field lines are deformed in the moving direction, which results in a similar phenomenon as the skin effect.

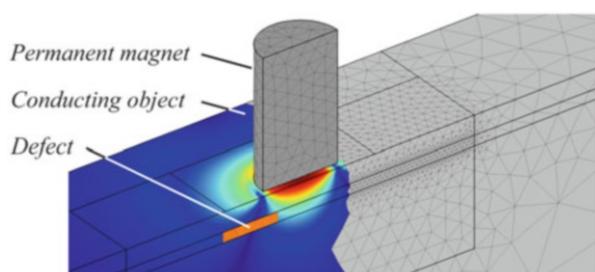
## Comparison of ECT and LET

A comparison between the ECT technique and LET is reported in (Carlstedt et al. 2013, 2014). In order to compare both methods from the numerical point of view, a detailed model of the applied ECT sensor is necessary. The probe under investigation was a differential-type probe PKA-48 (Rohmann GmbH), including secondary pickup coils. It was used with the ECT device Elotest N300 (Rohmann GmbH). Knowing the inner structure of the commercial ECT probe, finite-element models of both probes have been defined and used for the numerical comparison. First this analysis enabled the possibility to compare the defect response signals obtained experimentally with numerical simulations (Porzig et al. 2014). An example of 3-D FEM simulations is shown in Fig. 4.

Using the  $\mathbf{A} - \varphi$  potential formulation, the LET field problem can be described by (3), but without the external current density on the right-hand side. This formulation separates the two induction phenomena into the moving part  $\mathbf{v} \times \mathbf{B}$  and the time-changing part on the right-hand side. Depending on the definition of the frame of reference, two equivalent types of the general magnetic field induction equation can be distinguished (Zec 2013; Zec et al. 2013). In the so-called moving frame of reference, the global coordinate system is associated with the moving permanent magnet, i.e., the conducting object moves in the direction along the  $x$ -axis with velocity  $v$ . If the conducting object moves with a constant velocity and has a constant cross section normal to the direction of motion, e.g., the object is free of defects, the time derivative  $\partial\mathbf{A}/\partial t$  vanishes and (3) is reduced to a quasi-static approach.

The major differences between ECT and LET are given by the shape and the magnitude of the induced eddy current profile as well as by the evaluated signal. The impedance variations of the imaginary part  $\Delta Z_i$  and the back-induced voltage  $U_2$  in the secondary pickup coil from ECT have been compared to the force perturbations

**Fig. 4** 3-D FEM model used for LET simulations



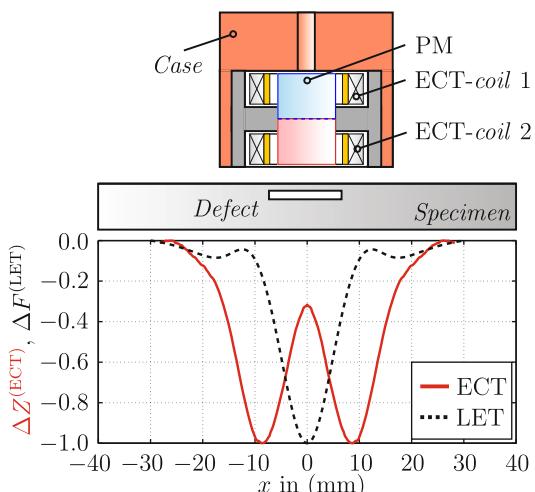
in the case of LET (Carlstedt et al. 2014). The normalized force perturbations (LET) and impedance perturbations (ECT) representing the normalized defect response signals one will get from both methods are shown in Fig. 5. The graph shows normalized signals of the drag force  $F_x$  together with the imaginary part of the secondary coil impedance at comparable source dimensions.

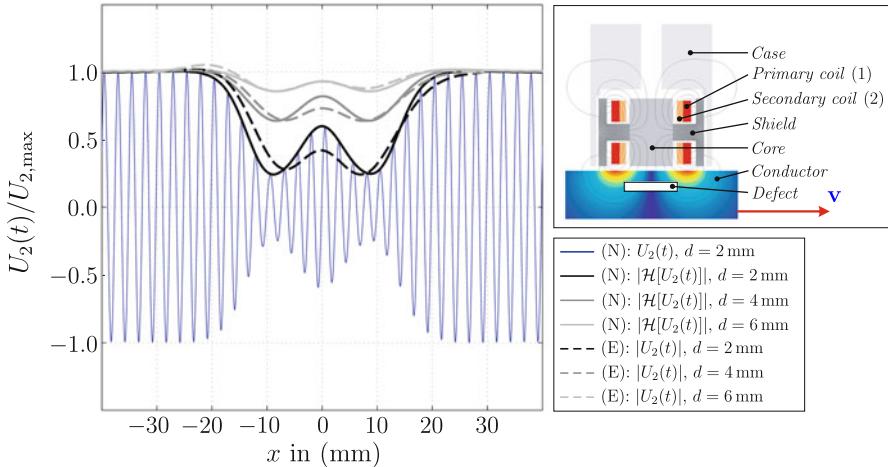
Usually, the ECT method is applied in stationary applications. However, when the object under test is moving relative to the ECT probe, the induced voltage in the pickup coil is modulated in the defect region. This effect is shown in Fig. 6.

If the velocity-to-frequency ratio  $v/f$  increases, the amount of sinusoidal periods in the defect region decreases. It is shown in (Carlstedt et al. 2014) that the use of the Hilbert transform of the secondary induced voltage  $\mathcal{H}[U_2(t)]$  is suitable to post-process the modulated defect response signals to determine the envelope of the modulated signal. In practice, this requires additional adjustments of currently available ECT devices. A direct comparison between both methods in terms of defect depth and velocity showed that with ECT, it was possible to detect defects of size  $[X_d, Y_d, Z_d] = [12 \text{ mm}, 2 \text{ mm}, 2 \text{ mm}]$  up to a depth of 6 mm at a velocity of  $v = 0.25 \text{ m/s}$  considering a frequency of  $f = 100 \text{ Hz}$ . In contrast, the LET method was able to resolve the defect up to a depth of 8 mm at  $v = 0.5 \text{ m/s}$ .

In many experiments, it was observed that the detection of subsurface defects in stacked aluminum sheets is possible for both testing techniques using the described experimental setup (Carlstedt et al. 2013). In the ECT method, the detection of a subsurface defect is mainly limited by the frequency-dependent penetration depth, i.e., if deep internal defects should be detected, the testing frequency has to be as low as possible. On the other hand, if low frequencies are used, the performance of the electronic amplifier becomes more important due to the weak signals. Furthermore, the testing speed is strongly restricted depending on the properties of defects, e.g., characteristic length and shape. In LET, a relative movement between the permanent magnet and the specimen is required to induce eddy currents. To create a sufficiently

**Fig. 5** Normalized defect response signals in the case of ECT and LET assuming equivalent dimensions (Carlstedt et al. 2013)





**Fig. 6** Modulated secondary induced voltage  $U_2(t)$  in the pickup coils in the case of moving objects under test ( $v = 0.25$  m/s,  $\sigma_0 = 30.61$  MS/m) (Carlstedt et al. 2014)

large Lorentz force, the relative velocity has to be high enough to detect small perturbations induced by subsurface defects. With the increasing speed, the absolute force and the force perturbations increase linearly at magnetic Reynolds number  $R_m < 1$ .

It can be concluded that both methods possess individual advantages. The classical ECT method is suitable to inspect stationary objects, which is not possible with LET or any other MIECT-type method. However, if the object is in motion, the use of alternating currents can be omitted. In this way, it is possible to apply permanent magnets, which produce considerably higher magnetic flux densities compared to current-carrying coils. Therefore, the magnitude of the desired force signal is theoretically adjustable with the velocity for optimal utilization of the applied force sensor. In practice, the force sensor is sensitive to unwanted vibrations of the environment and the system itself.

To summarize, both testing techniques are highly dependent on the used sensors and measurement electronics as well as on the available testing speed. Consequently, many areas of application of LET and ECT, respectively, will be different. Comparative studies (Carlstedt et al. 2013, 2014) showed that LET is a promising and competitive alternative to traditional ECT methods considering the contactless evaluation of moving electrical conductors.

## Motion-Induced Eddy Current Testing (MIECT)

### Numerical Simulations

The intrinsic phenomena associated with electric and magnetic fields affect almost all aspects of our everyday life. This is accompanied by continuous development and design of more sophisticated electromechanical devices, which result in better

functionality, higher efficiency, and increased safety. Depending on the particular application, this task strongly relies on the accurate modeling of the electromagnetic fields within the device. Apart from providing exact and fast solutions, a closed-form analytical expression helps in better understanding of the underlying physical phenomena associated with the problem under investigation. Unfortunately, these solutions are not always available, and they can be obtained only for some simplified device and field configurations. Thus, in NDT&E applications, the development and optimization of various testing techniques are performed using numerical methods (Ida 1995). Due to its ability to handle complex geometries, anisotropic and inhomogeneous material properties, a widely used numerical method in NDT&E applications, are the finite element method (FEM) (Ida and Bastos 1997). In general, analysis of LET systems requires accurate and time-efficient numerical approaches to allow either extensive scans of an object under test or parametric and optimization studies. For the implementation of the proposed approaches, the commercial software package COMSOL Multiphysics (COMSOL 2018) has been used.

## Computation of Eddy Current Distributions in Moving Parts

The LET belongs to a special class of electromagnetic field phenomena in which various effects caused by parts set in a relative motion occur. These effects represent the basic operating principle over a wide range of electromechanical devices in different application areas such as electrical machines, magnetic levitation systems, inductive heating, eddy current brakes, nondestructive testing and evaluation (NDT&E), etc. Due to its tremendous industrial relevance, the application of FEM to this particular type of field problems, also known as the moving eddy current problems, has undergone extensive research over the past decades (Ying et al. 2007).

In principle, independently of the actual type of motion (translation or rotation), all existing techniques for simulation of general moving eddy current problems can be classified into (i) fixed grid methods and (ii) moving or time-changing grid methods (Demenko 1996).

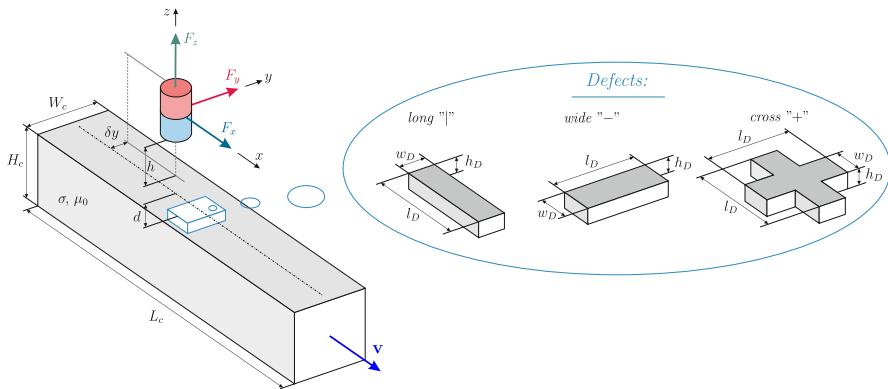
The fixed grid methods are usually applied to 2-D/3-D static or time-harmonic eddy current problems involving uniformly moving conducting parts having invariant cross section in the motion direction (Binns et al. 1992; Bird and Lipo 2009; Demenko 1996; Rodger and Eastham 1985; Rodger et al. 1990, 1991; Ying et al. 2007; Ziolkowski and Brauer 2010). The field problem is formulated in the moving frame of reference where the additional velocity term ( $\mathbf{v} \times \nabla \times \mathbf{A}$ ) is used to describe the contribution of the induced eddy currents within the moving electrically conducting part. This approach, also referred to as the quasi-static approach (QSA) (Ziolkowski and Brauer 2010), is very efficient in terms of computational time since only one stationary analysis needs to be performed to obtain an accurate steady-state solution. For some simple device configurations, several authors combine FEM with analytical solutions as well (Bird and Lipo 2009; Kirpo et al. 2011) and report a considerable reduction of the simulation time and increased accuracy. However, apart from simple geometries, the analysis is restricted to stationary and time-harmonic problems.

The moving grid methods are more general, and they can be applied to simulate a wide variety of electromechanical devices involving linear or rotational movement. In principle, from the model topology point of view, all available techniques are quite similar. The main idea is to decompose the whole computational domain into two parts associating them with the moving or with the fixed part of the assembly (Demenko 1996; Trowbridge and Sykulski 2006; Ying et al. 2007). Within each part, the governing equations are solved in their own frame of references, whereas the relative displacement and the field coupling are provided on the introduced interface (Biddlecombe et al. 1998; Davat et al. 1985). Depending on the actual interface, which can have constant or variable lift-off distance, to achieve the coupling, many different techniques have been applied.

One of the most widely applied moving grid methods to model relative displacements in general eddy current problems is the so-called sliding mesh technique (SMT) (Buffa et al. 2000). Similarly to all other methods, SMT, also referred to as the moving mesh method (Ying et al. 2007) or slip surface method (Preston et al. 1988), requires two independent meshes to be defined. To provide the relative displacement, the meshes are simply slid relatively to each other eliminating any need to alter their structure. The governing equations are solved independently in the fixed reference frame of each moving part, thereby avoiding the convection (velocity) terms. Depending on the mesh distribution along the introduced interface, which can be conforming or nonconforming, the field continuity can be preserved using several coupling techniques. In the case of conforming meshes, the unknown potentials on each side of the sliding interface are made equal in the same way as Dirichlet boundary conditions are imposed (Muramatsu et al. 1996, 1999; Preston et al. 1988; Yamazaki 1997, 1999; Ying et al. 2007). However, the displacement is strictly controlled by the size of the finite elements in the motion direction and the time-step size of the transient solver. To overcome this limitation, nonconforming meshes along with Lagrange multipliers have been introduced (Rodger et al. 1990). The Lagrange multiplier approach introduces an additional set of variables on the sliding interface which ensures the continuity of the field in a weak sense (Golovanov et al. 1998; Lai et al. 1991; Leonard et al. 1993; Marechal et al. 1992). Unfortunately, the existence of additional variables considerably deteriorates the conditioning of the stiffness matrix (Antunes et al. 2006a, b; Golovanov et al. 1998).

## Numerical Modeling

The main aim of this section is to introduce a new FEM-based methodology, which can be used to analyze and develop future LET systems. The particular emphasis is placed on the reduction of the overall computational requirements while maintaining the accuracy of the solution. Additional goals include development of simplified numerical models which enable fast 2-D and 3-D LET analysis in conjunction with the verification of assumed simplifications. The problem is simplified step-by-step, starting from time-dependent approaches, applying quasi-static approximations, and assuming a weak reaction from the conductor. For comparison and verifications of



**Fig. 7** Definition of the LET benchmark problem. The conductor contains three types of defects: long ("|"), wide ("—"), and cross ("+"')

different approaches, a benchmark problem which represents a typical LET configuration has been considered.

### Benchmark Problem

As a LET benchmark problem, a generic conductor with pre-defined artificial defects, moving across the static magnetic field, is chosen (Fig. 7). The conductor under test is considered to be nonmagnetic with the electrical conductivity denoted by  $\sigma$  and magnetic permeability equal to the permeability of vacuum  $\mu = \mu_0$ . It has a rectangular cross section determined by its width  $W_c$  and height  $H_c$ , whereas its length is denoted by  $L_c$ . A cylindrical permanent magnet described by magnetization  $\mathbf{M}$  is used as a source of the static (primary) magnetic field. The diameter and the height of the magnet are denoted by  $D_m$  and  $H_m$ , respectively. The magnet is placed centrally above the conductor under test ( $\delta y = 0$ ) at a lift-off distance  $h$ .

For the analysis, three different types of artificial defects, namely, long ("|"), wide ("—"), and cross ("+"'), have been considered. Defects are placed centrally within the conductor at depth  $d$  below its surface. They are characterized by their width  $w_D$ , height  $h_D$ , and length  $l_D$ . The conductivity of the defect is denoted by  $\sigma_d$ . In order to reduce the number of dependent variables, the magnetic Reynolds number ( $R_m$ ) has been used for the analysis (4). It involves the characteristic length-scale parameter  $L$  whose definition depends on the particular problem at hand and characterizes the moving conductor.

### Logical Expression Approaches

The main idea of the logical expression approaches (LEA) that allows fast computations of 2-D/3-D eddy current problems including parts in relative motion is presented (Zec et al. 2013). Using the proposed methodology, the spatial coordinates of moving parts, either conducting or nonconducting, are modeled on a fixed computational grid using logical expressions (LE). By applying the principles of Boolean algebra directly in finite element analysis (FEA), the shape of moving parts is determined on the fly by calculating the constraints given by LE and filtering the

finite elements in those domains where LE are introduced. Figure 8 shows three basic geometrical primitives, i.e., box, cylinder, and sphere, which are modeled using LE. Independent to the type of reference frame used in LET analysis, modeling of moving parts using LE requires the existence of a homogeneous zone in which these expressions are applied. This zone is referred to as the moving domain, and it is determined by the shape of the moving part and its relative displacement  $L$  (Fig. 8). In order to introduce the motion and to determine the shape of moving parts, the constraints given by LE are defined as time dependent. This step represents the basic idea of the logical expression approach (Zec et al. 2013).

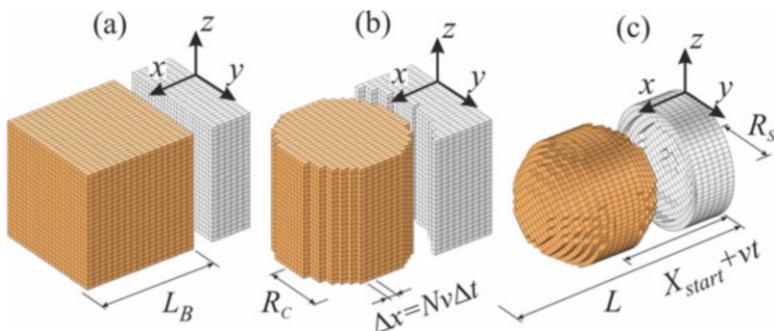
Depending on the definition of the global frame of reference, the LET analysis has been performed using two different implementations of LEA:

1. In the case of fixed frame of reference, the logical expressions are used to model the motion of the permanent magnet.
2. In the case of moving frame reference, the logical expressions are used to model the relative motion of the defect.

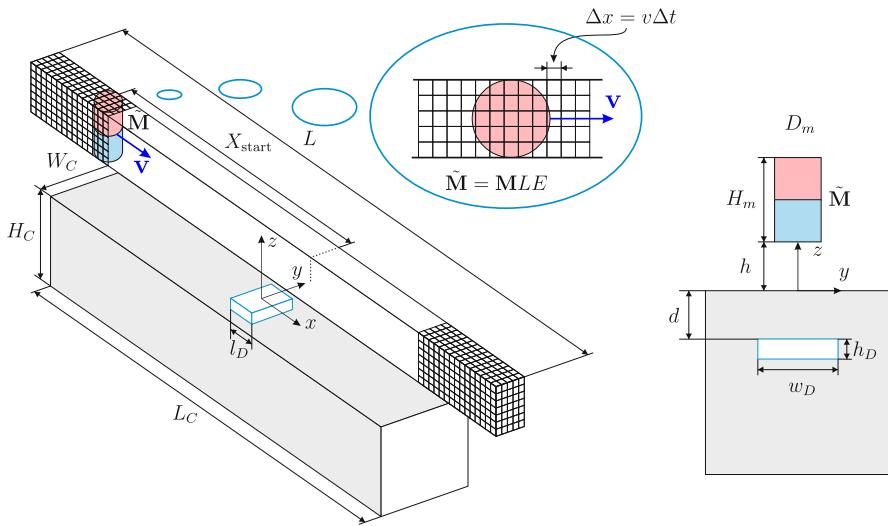
These two specific LEA implementations are referred to as the moving magnet approach (MMA) and the moving defect approach (MDA), respectively.

### Moving Magnet Approach

In the implementation of the moving magnet approach (MMA), the global coordinate system is associated with the conducting object (fixed frame of reference), and the logical expressions are used to describe the motion of the used cylindrical permanent magnet (Fig. 9). The magnet is moving with constant velocity  $v$  along the model  $x$ -axis in close vicinity of the conductor containing an artificial defect below its surface. The moving domain is defined in the surrounding air region. The cross section of the moving domain is determined by the height  $H_m$  and the diameter  $D_m$  of the magnet, while its length depends on the starting position  $X_{\text{start}}$  and its relative displacement  $L$ . In general, the starting position  $X_{\text{start}}$  is the distance of the moving object to the origin of the coordinate system at  $t = 0$ . It has to be large



**Fig. 8** Basic geometric primitives defined by logical expressions: (a) box, (b) cylinder, and (c) sphere



**Fig. 9** Implementation of the moving magnet approach (MMA). The moving domain is defined in the air region

enough to avoid any influence of the used initial conditions on the resulting Lorentz force perturbations. Applying the modified vector potential  $\mathbf{A}^*$  (Zec 2013), the governing system of equations in MMA is given by:

$$\nabla \times \left( \frac{1}{\mu_0} \nabla \times \mathbf{A}^* - \mathbf{M} \right) = -[\sigma] \frac{\partial \mathbf{A}^*}{\partial t} \quad (5)$$

$$\nabla \cdot \left( -[\sigma] \frac{\partial \mathbf{A}^*}{\partial t} \right) = 0 \quad (6)$$

where  $[\sigma]$  is a diagonal tensor of electrical conductivity  $[\sigma] = \text{diag}(\sigma_{xx}, \sigma_{yy}, \sigma_{zz})$ . This requires both nodal and edge finite element formulations in single computational domain. Due to the fact that (5) and (6) do not introduce the additional velocity term, the resulting system of equations remains symmetric. Another important feature of MMA is that the stiffness matrix has to be assembled only once during the entire motion of the magnet which additionally reduces the total computational time. This is because the motion is provided by simple modification of the magnetization vector  $\mathbf{M}$  which appears as a source term in the resulting FEM formulation.

### Moving Defect Approach

In the moving defect approach (MDA), the global coordinate system is assigned to the permanent magnet (moving frame of reference). In this reference frame, the magnet is stationary, and the conductor is moving in opposite direction with velocity  $-v$ . For the given LET problem, the use of LE to describe the motion of the whole conducting domain would be computationally very expensive. However, the implementation of LEA in the

moving reference frame can be considerably simplified if only perturbations of Lorentz force caused by defects are required, which is in fact a typical LET problem. Thus, it is sufficient to model only the movement of the defect relatively to the magnet, instead of modeling the motion of the whole conductor.

In MDA the moving domain is defined entirely inside the conductor where LE are used to model the motion of the particular defect (Zec et al. 2013). The shape and position of the moving domain are defined by the cross section of the defect  $w_D \times h_D$ , its relative displacement  $L$ , and depth  $d$ . Similar to the previous MMA, applying the modified vector potential  $\mathbf{A}^*$  in MDA results in the system of governing equations given by:

$$\nabla \times \left( \frac{1}{\mu_0} \nabla \times \mathbf{A}^* - \mathbf{M} \right) = [\sigma] \left( -\frac{\partial \mathbf{A}^*}{\partial t} + \mathbf{v} + \nabla \times \mathbf{A}^* \right) \quad (7)$$

$$\nabla \cdot \left[ [\sigma] \left( -\frac{\partial \mathbf{A}^*}{\partial t} + \mathbf{v} \times \nabla \times \mathbf{A}^* \right) \right] = 0 \quad (8)$$

This requires both nodal and edge finite element formulations to be applied in the computational domain. In contrast to MMA, the formulation used in MDA involves the additional velocity term which makes the resulting system of equations non-symmetric. Additionally, the modification of the electrical conductivity by time-dependent LE, introduced by MDA, modifies the resulting stiffness matrix as well. Therefore, the stiffness matrix has to be reassembled at every time step which increases the computational time compared to MMA.

### Quasi-static Approach

The moving magnet approach (MMA) and the moving defect approach (MDA) assume no simplifications for the given LET analysis. They offer accurate results for any relative testing velocity  $\mathbf{v}$  between the magnet system and the conductor and for any material and geometry parameters involved. Thus, they are valid for finite values of the magnetic Reynolds number, whether the conductor contains material defects or not. If the LET configuration under investigation is time independent, i.e., it involves uniformly moving conductors with a constant cross section normal to the direction of motion (conductors free of defects), the analysis can be considerably simplified. A single stationary analysis can be performed to obtain an accurate steady-state solution, e.g., the Lorentz force acting on the magnet system. This assumption requires the moving frame of reference, where the additional velocity term ( $\mathbf{v} \times \nabla \times \mathbf{A}$ ) is used as a source of the induced eddy currents inside the conductor in uniform motion

$$\nabla \times \left( \frac{1}{\mu_0} \nabla \times \mathbf{A} - \mathbf{M} \right) = [\sigma](-\nabla \varphi + \mathbf{v} + \nabla \times \mathbf{A}) \quad (9)$$

$$\nabla \cdot [[\sigma](-\nabla \varphi + \mathbf{v} + \nabla \times \mathbf{A})] = 0. \quad (10)$$

The second equation results from the current conservation law  $\nabla \cdot \mathbf{J} = 0$ , and it is an additional equation for the electric scalar potential  $\varphi$ . This system of equations

takes the deformation of the magnetic field lines correctly into account making it valid for any value of the  $R_m$ . Although (9) and (10) provide fully correct solutions only for conductors without any material defects, their use can be still extended to NDT applications as well. It has been shown that for LET systems resulting in small  $R_m$ , they can be used for fast Lorentz force calculations on the moving magnet even for conductors with defects (Ziolkowski and Brauer 2010). The method is referred to as a quasi-static approach (QSA). As a direct consequence of low magnetic Reynolds numbers ( $R_m < 1$ ), the diffusion time of the magnetic field into the conducting object was estimated as  $\tau \sim R_m \cdot L/v$ , where  $L$  is the characteristic length scale of the conductor. This basically justifies the instantaneous field reaction ( $\partial\mathbf{B}/\partial t \rightarrow \mathbf{0}$ ) to any perturbation of induced currents, which is assumed in QSA. Nevertheless, if this is not the case, the full transient form of (9) and (10) has to be considered, representing the governing equation of already presented moving defect approach (MDA). In the implementation of QSA, only the change in relative position between the magnet and the defect must be provided. This is done either by moving the magnet system relative to the defect or vice versa (Ziolkowski and Brauer 2010). In any case, this requires a time-consuming re-meshing procedure of the entire model geometry for each new configuration. The re-meshing of the geometry can be avoided if the basic principle of the logical expression approach (LEA) is combined with the quasi-static formulation given by (9) and (10). In this LEA implementation, the time variable used in different LE is just a parameter which needs to be changed from one stationary solution to another providing the displacement of the moving part (magnet or defect). This means that the same geometry used for the implementation of the LEA can be used for implementation of QSA as well. The only difference introduced by QSA is in the governing equation in the conducting region, which is now in its stationary form and contains an additional scalar potential  $\varphi$ . In regions free of eddy currents (surrounding air region and permanent magnet), the magnetic scalar potential is used.

### **Weak Reaction Approach**

The induction problem at hand can be further simplified in the case of low magnetic Reynolds numbers ( $R_m \ll 1$ ). In this case, the induced eddy current density is so small that its magnetic field  $\mathbf{B}^{(s)}$  is vanishingly small compared to the primary magnetic field  $\mathbf{B}^{(p)}$  of the magnet system. By setting  $\mathbf{B}^{(s)} = \mathbf{0}$ , the magnetic and electric fields are decoupled and, therefore, can be treated independently. Hereinafter, this effect will be referred to as a *weak reaction* by the conductor to the magnetic field. Special attention must be paid to the emerging Lorentz forces when using weak reaction-based approaches. By neglecting the secondary magnetic field, the spatial symmetry of the electric and magnetic field is enforced. As a consequence, the lift component of the Lorentz force vanishes if the conductor is free of defects and if the magnet is far from any outer edge of the conductor, such that  $F_z^{(0)} = 0$ . However, in the presence of defects, the symmetry of the fields no longer holds, and the defect response signal  $\Delta F$  can be determined.

## Comparison of Numerical Approaches

The approaches discussed above differ in their treatment of the secondary magnetic field  $\mathbf{B}^{(s)}$ . Hence, it is necessary to investigate their applicability in terms of magnetic Reynolds number  $R_m$ . The ratio between primary and secondary magnetic fields depends on the underlying geometry of the problem, as indicated by the characteristic length  $L$  in  $R_m$ . To conduct an expressive comparison, an exemplary LET problem that corresponds to the dimensions of the available laboratory setup is defined (Zec 2013).

The absolute defect response signal (ADRS)  $\Delta\mathbf{F}$  is defined as the force perturbation resulting from a defect. Because Lorentz forces are also present in the unperturbed case, the ADRS can be mathematically defined by the difference between the perturbed force profile  $\mathbf{F}$  and unperturbed force profile  $\mathbf{F}^{(0)}$ :

$$\Delta\mathbf{F} = \mathbf{F} - \mathbf{F}^{(0)}. \quad (11)$$

Thus, the ADRSs are calculated for different magnetic Reynolds numbers  $R_m$  by varying the velocity.

No significant differences can be identified between time-dependent approaches and WRA, which indicates that time-dependent effects are negligible. When increasing  $R_m$ , secondary fields and time-dependent effects become prevalent, resulting in nonsymmetric field and force profiles. The ADRS obtained using WRA retains its symmetry because the secondary fields are neglected. Consequently, WRA overestimates the ADRS amplitude compared to time-dependent approaches in the case of high  $R_m$ . The ADRS obtained using QSA is closer to ADRS values obtained using MDA and MMA, because it includes the stationary part of the secondary magnetic field ( $\mathbf{B}^{(s)} \neq 0$ ). In the case of high  $R_m$ , the time-dependent part of the secondary magnetic field  $\partial\mathbf{B}^{(s)}/\partial t$  has an increasing influence on the ADRS. By comparing the results of QSA to those of MDA and MMA, a delayed and damped force response can be seen. As expected, the solutions from MMA and MDA are equivalent and yield very similar force profiles, because they only differ in the definition of the frame of reference. These effects pertain to both isotropic and anisotropic specimens. However, the ADRS has higher amplitudes in the anisotropic case than in the isotropic case. This phenomenon can be explained based on the imposed condition that  $J_z = 0$  because  $\sigma_{zz} = 0$ . As a consequence, the current flows around the defect only in the  $x - y$ -plane (i.e., not vertically). This phenomenon positively influences the resulting Lorentz force in terms of the ADRS amplitude. The shape of the ADRS is weakly influenced by this condition because in the unperturbed case, the induced eddy currents already flow solely in the  $x - y$ -plane. However, some differences between the two cases can be identified. Their anisotropic profiles show slightly sharper ADRSs, producing higher gradients. It can be concluded that the anisotropy condition influences the profile but does not significantly change it. This result confirms the applicability of layered specimens for the investigation of deep-lying defects. WRA and QSA are on nearly the same level, up to moderate values of  $R_m$ . However, the error in WRA increases significantly when  $R_m$  reaches values of

roughly 10, which corresponds to velocities of roughly 6 m/s or 10 m/s for specimens made of copper or aluminum, respectively.

One major drawback of WRA is the absence of the unperturbed lift force  $F_z^{(0)}$ , which is an immediate result of the decoupling of electric and magnetic fields. The imposed symmetry in  $\mathbf{B}^{(P)}$ , and therefore also in  $\mathbf{J}$ , eliminates the lift force after the volume integration. Apart from testing the applicability of different numerical approaches for LET problems, the numerical studies also provide deeper insights into the underlying physics of motion-induced eddy currents.

---

## Measurement and Sensor System

### Force Measurement

There is a wide range of techniques available to measure forces. The most common measurement principles of force transducers are introduced with special focus on strain gauge load cells and piezoelectric crystal force transducers. Furthermore, the characteristics of force measurement systems and the importance of calibration have to be considered.

The force is a physical vector quantity that acts on a single point. In order to measure this vector, it has to be translated into a scalar quantity. Since real materials have limited permissible stress, a force cannot be transferred via a single point but always by a finite surface. So strictly speaking, not the force itself is measured but the stress tensor field caused by the force.

A real force measurement system is therefore composed of a force transducer and the associated instrumentation, as well as perhaps mechanical installation aids. A force transducer is a device which converts the applied force into a measurable scalar quantity, e.g., change of electric resistance, through a known physical relationship. The instrumentation associated with a force transducer is used to generate an analogue or digital electrical output to represent the indicated value. Depending on the requirements of the measurement application, the instrumentation may contain a number of separate elements for signal conditioning, indication, analogue-to-digital conversion, and data collection.

Force measurement systems are based on different physical principles but can be described by several common characteristics. The behavior of all these systems can be expressed by plotting the response curve, which represents the indicated output value against the applied force. An ideal response curve is a straight line from zero to the rated capacity of the force measurement system and then back again to zero. Real measurement systems differ from this ideal curve in multiple ways. These are commonly categorized by their systematic deviation with respect to the least squares optimal line through the origin for increasing outputs.

An important concept in measurement systems is known as repeatability which is described by the agreement of the outputs for repeated applications of the same load. In practice, the repeatability of transducers is also provided by many manufacturers.

However, this value can only be a reference for optimal conditions during the testing procedure. In operation it is generally an overestimation of the performance of the force measurement system due to limitations of the actual electrical measurement equipment, present temperature gradients from first to last application, and other environmental influence quantities.

Further characteristics of force measurement systems can be summarized by the imperfections of applying the force to be measured to the loading surface of the transducer. One of the most important specifications is thereby the sensitivity to off-axis forces which result in parasitic torque. For single-component transducers, it is also important to consider the sensitivity to orthogonal forces which is equivalent to misalignment of the transducers' principal axis to the force to be measured.

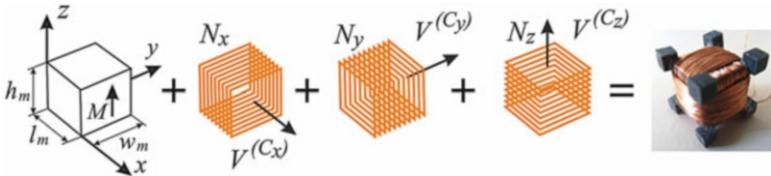
## Differential Lorentz Force Eddy Current Testing Sensor

Based on the LET analysis presented previously, it might be necessary to perform the testing with high velocity, thus using permanent magnet configurations that are better adapted to the current measurement task. Apart from increasing the force perturbations, this would increase the absolute values of the Lorentz force, due to Faraday's law of induction. This could even be a few orders of magnitude larger than the perturbations caused by defects.

Unfortunately, precise measurements of small force variations in a relatively large range of applied forces are very difficult. Thus, there is a strong demand for the usage of differential force measurements in the LET system. Similar tendency is observed in traditional ECT systems as well. In ECT, various differential pickup probe configurations offering higher testing sensitivity have already been designed and successfully implemented (Chari et al. 1990; Mook et al. 2011). Currently, there are several possibilities to obtain the differential force signals resulting from LET system. However, the application of commercially available differential force sensors would lead to higher spatial integration requirements and considerably higher costs. This would be even more important when designing sensor arrays for LET, which could simplify and advance the defect detection and reconstruction, respectively.

Further, a simple and low-cost modification of LET setup is presented which could be used for measurements of differential Lorentz force signals caused by material defects. The proposed modification affects the magnet used where three independent and passive pickup coils have been wound on its surface (Fig. 10). The principal idea is to use voltages induced in the additional coils and correlate the voltage signals with the corresponding differential Lorentz force signals exerted on the magnet. In fact, the resulting magnet system can be applied directly to the existing LET experimental setup, i.e., it can be used as a complete differential Lorentz force sensor. Since this specific modification of the LET system allows differential Lorentz force measurements, the proposed technique has been termed as differential Lorentz force eddy current testing (DiLET).

The concept of the corresponding sensor is based on (Uhlig 2014; Zec 2013; Zec et al. 2015), where a set of passive coils fixed to a PM is proposed to allow the



**Fig. 10** DiLET – scheme of 3-D differential Lorentz force sensor

detection of perturbations in the eddy current distribution caused by defects inside the specimen.

The main idea of using a simple coil system to obtain the differential Lorentz force signals can be traced back to the following force relation (Haus and Melcher 1989; Ramos and Lopes Ribeiro 2014).

$$\mathbf{F} = \int_{\Omega_M} (\mathbf{M} \cdot \nabla) \mathbf{B} d\Omega = - \int_{\Omega_C} \mathbf{j} \times \mathbf{B} d\Omega \quad (12)$$

The primary magnetic field  $\mathbf{B}^{(p)}$  produced by a permanent magnet is constant in time, while the secondary magnetic field  $\mathbf{B}^{(s)}$ , connected with the eddy current distribution inside the specimen, is time dependent when a defect is present. Thus, the induced voltage  $V_i, i \in \{x, y, z\}$  in a coil fixed to the PM is proportional to perturbations of the secondary magnetic field  $B_i^{(s)}$  and therefore sensitive to disturbances caused by the defect. As shown in (Zec 2013), the induced voltage  $V_i$  is proportional to the time derivative of the force component parallel to the respective coil axis.

## Optimization of the Permanent Magnet System

Comparative studies between LET and ECT indicated the potential and competitiveness of LET (Carlstedt et al. 2013, 2014). However, the performance of an LET system can be enhanced further by applying optimization schemes to determine advanced magnet systems with improved characteristics.

The optimization goal is to maximize the response resulting from an inclusion surrounded by conductive material, thereby increasing the signal-to-noise ratio and, hence, improving the detection rate. However, due to the large variety of NDT problems, it is self-evident that the final details of an optimized setup strongly depend on the detection goal and external testing conditions for the particular application.

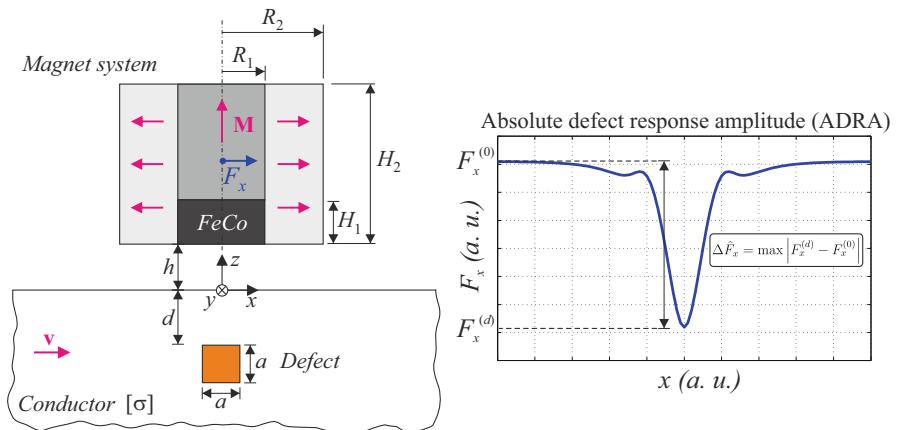
The optimal magnet design is taking into account the tasks to evaluate non-magnetic, electrically conducting specimens. However, it is also applicable to ferromagnetic specimen if the objective function evaluation is adapted to the corresponding requirements. The optimization is performed under the assumptions of a smooth specimen surface and that the defect is located far from any lateral

boundary to neglect parasitic edge effects. Since the resulting Lorentz force profile depends on the shape and the depth of the inclusion, an equivalent defect of cuboidal shape is defined to represent a general flaw. The assumptions can be modified to any particular case of interest, since this would involve only the geometry of the specimens defined in the forward solution, which are described in one of the following sections. The optimization is performed with respect to the drag force  $F_x$  and the associated absolute defect response amplitude (ADRA)  $\Delta\hat{F}_x$ , resulting from the difference between the unperturbed drag force  $F_x^{(0)}$  and the perturbed drag force  $F_x^{(d)}$ . The force profile is symmetric if the interaction between the primary magnetic field  $\mathbf{B}^{(p)}$ , generated by the permanent magnet, and the secondary magnetic field  $\mathbf{B}^{(s)}$ , from the induced eddy currents, is negligible.

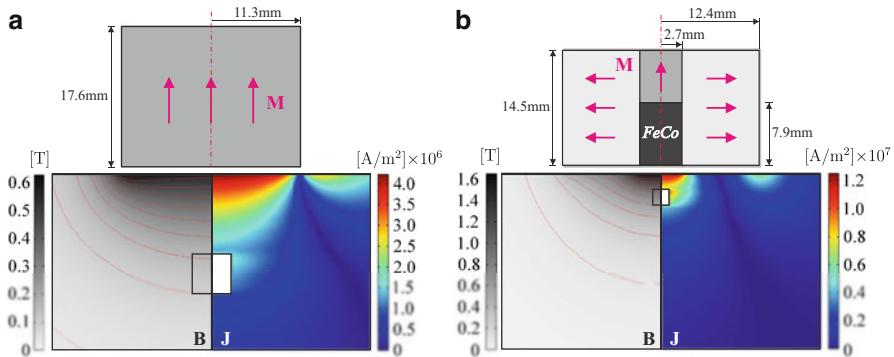
Figure 11 shows the geometrical parameters of the problem together with the ADRA. The specimen is modeled as a pseudo-infinite half-space including a defect with edge length  $a$ , located at a depth  $d$ . The magnet system is located at a lift-off distance  $h$  above the specimen. The optimization scheme covers but is not limited to isotropic specimens ( $\sigma_{xx} = \sigma_{yy} = \sigma_{zz}$ ) and laminated structures ( $\sigma_{xx} = \sigma_{yy} \neq 0$ ,  $\sigma_{zz} = 0$ ).

Two distinct magnet systems of both scenarios are compared, considering the same magnet volume  $V_m$ . For medium-sized deep defects, a cylindrical magnet with a volume ratio of  $V_m/V_d = 56$  is chosen to be optimal. The Halbach structure magnet system optimized for small subsurface defects has a volume ratio of  $V_m/V_d = 875$ . The geometrical parameters are obtained by de-normalizing both systems assuming a lift-off distance and hence geometric scale of  $h = 1$  mm. The spatial distribution of the magnitude of the magnetic flux density  $\mathbf{B}$  and the induced eddy current density  $\mathbf{J}$  is shown in Fig. 12.

The eddy currents for regular cylindrical magnets (Fig. 12a) are less concentrated compared to Halbach systems (Fig. 12b). The Halbach structure leads to a



**Fig. 11** Parameters of the LET setup, design variables of the magnet system, and depth and resolution, which can be controlled by means of the diameter of the exciting coil (Mook et al. 2007)



**Fig. 12** Magnitude of the magnetic flux density  $\mathbf{B}$  and induced eddy current density  $\mathbf{J}$  of the optimized magnet systems for the case of anisotropic specimens, assuming a lift-off distance of  $h = 1$  mm. Both generate an unperturbed drag force of  $F_x^{(\text{lab})} = 3$  N. Cross section of (a) a cylindrical magnet and (b) Halbach structure

considerably more focused magnetic flux and eddy current distribution under the inner part of the magnet system. The flux density is increased to 1.6 T on the surface of the specimen, which is significantly larger compared to standard magnet systems (Weise et al. 2015a).

When considering high-speed applications, secondary magnetic fields become prevalent and cannot be neglected. In this case, the magnetic field formulation used in the forward solver can be adjusted (e.g., to an  $\mathbf{A}, \varphi - \psi$  formulation), albeit at the expense of computational cost. In ECT, there exists a trade-off between penetration the absolute defect response amplitude used as objective function (Weise et al. 2015a)

This fact can be also observed in LET and is inherently a limiting factor of both methods. The results demonstrate the advantage of combining active and passive magnetic materials in the form of a Halbach structure in an LET sensor for selected applications. The results of the unconstrained optimization demonstrate that the use of those structures is counteracting the trade-off between penetration depth and resolution, revealing additional potential of future sensor systems. In the present case, the optimization procedure has been performed considering two different defect scenarios. The associated optimal magnet designs are manufactured and made available for experimental studies presented in (Carlstedt 2016; Mengelkamp 2016). A defect depth study, adopting the concept of a quasi-infinite crack, revealed a current detection depth limit of about 35 mm.

## Lorentz Force Evaluation

Nondestructive material testing and evaluation is a vast interdisciplinary field as well as a challenge due to the variety of applications. Whereas the focus of nondestructive testing (NDT) is to detect and localize anomalies within a

specimen, the reconstruction of defect properties (dimensions, shape, structure, composition) and their influence on the material's usability is the focus of nondestructive evaluation (NDE).

Defect identification and assessment are very important aspects of quality assurance. Nondestructive material testing is understood as the noninvasive examination of any type of specimen without changing or altering the properties of the body under test to check whether the specimen contains anomalies. Anomalies are any type of defect or change in the material properties that can be of natural or artificial origin, influencing the usefulness or serviceability of that object (Hellier 2013). Nondestructive testing has turned from a rather empirical procedure dependent on the experience of the examiners into a more quantitative measurement technique that serves to determine the influence of material anomalies on the structural health of the object (Achenbach 2000).

Biomedical applications often require the determination of electrical conductivity of human tissues. It is a common approach to use induction coils inducing electrical currents in the tissues. This results in changes of the coil's impedances. These changes are used to obtain information about the tissue conductivity. Similarities and dissimilarities between appropriate sensor setups have been investigated and compared for biomedical applications (Petković 2013). A very similar approach can be used to determine the conductivity distribution of metallic specimen. If the permanent magnet system is approximated by a dipole model, the same strategy for the solution of the inverse problem can be applied. This serves as the basic idea for proposing a new method for noncontact, nondestructive evaluation of solid conductive materials, termed Lorentz force evaluation (LFE). In contrast to the bioelectromagnetic application, where magnetic flux is measured at points above the conductor, in LFE the Lorentz force acting on a permanent magnet moving relative to the specimen is measured.

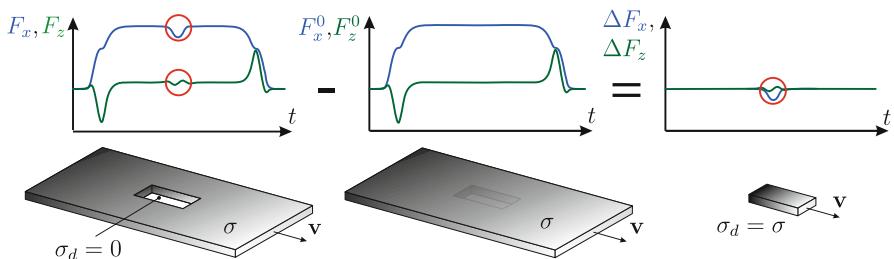
The inverse problem of the LFE technique, which has been introduced for inverse calculations of LET signals in (Petković et al. 2013), is to characterize the geometry of the underlying defects. A straightforward approach is applied to solve this inverse problem directly and estimate the defect parameters. The geometrical parameters of the defect, such as the center of gravity, depth, and extension, can be defined as the unknown variables. Other approaches are to determine the eddy current distribution and the conductivity distribution in the conductor. In these reconstruction approaches, the solution vector can compose the unknown moments of the equivalent current dipoles or the unknown voxel conductivities. Solving an inverse problem implies the minimization of the error between measured data and forward calculated data with respect to the unknown parameters.

Material anomalies, such as changes in conductivity, defects, cracks, or inclusions, distort the eddy current distribution in the object under test and, consequently, also the Lorentz force measured at the magnetic system. Thus, defects/anomalies in the conductive material produce perturbations in the Lorentz force signals. Extensive basic research has been done to show that the direct relationship between changes in force and material anomalies can be used to detect defects. If the detected defect has to be identified with respect to depth, material properties,

and shape reconstruction, the procedure is called motion-induced eddy current testing and evaluation (MIECTE).

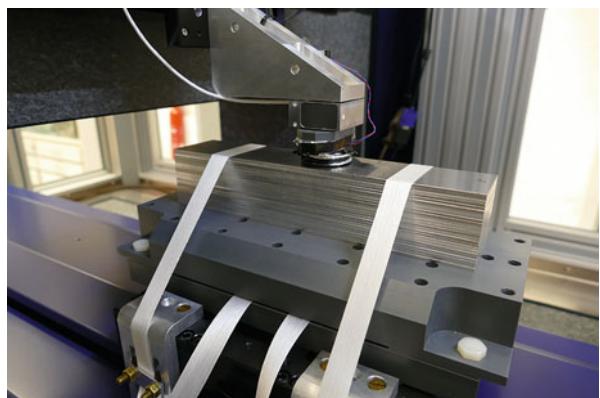
In general, the defect response signal (DRS) defined as the difference between the Lorentz force determined for the defective object and the defect free object is used in MIECTE. In the simplest approach (Petković et al. 2013), the DRS can be estimated using a set of current dipoles located only in the defect region (Fig. 13). The current dipoles are calculated from the distribution of eddy currents determined for the defect-free conductor. However, the DRS found in this way is far from the actual signal because it does not take into account the disturbance of the flow of eddy currents around the defect. DRS quality can be improved by the use of the so-called extended area approach (EAA) in which the influence of disturbed eddy currents flowing around the defect is taken into account (Ziółkowski 2015).

In the LFE, minimum norm estimation methods are often applied to the reconstruction of eddy current distributions, which are responsible for defect response signals that are measured by the LET system (Fig. 14). Solving an inverse problem in a fully three-dimensional source space requires the use of high computational resources, including large memory space and long computation times. In order to avoid the inversion of very large kernel matrices, the defect reconstruction procedure can be divided into three steps: (1) determination of a depth of the intermediate plane



**Fig. 13** Simplest approach for modeling a defect response signal from a defect in a conductive specimen

**Fig. 14** Experimental setup for the LET measurements in the laboratory



of a defect, (2) determination of the length of a defect in a moving direction of a specimen, and (3) reconstruction in the defect midplane parallel to the upper surface of the bar and the scanning plane. Determination of the depth of long defects can be precisely determined using both measurement and data obtained by finite element method (Petković et al. 2013). In the case of wide defects, a depth of its intermediate plane is slightly moved toward the surface but still being in an acceptable range of error (i.e., less than 1 mm). Determination of a subsurface defect depth involves as much as possible a priori knowledge about the defect, e.g., it is usually assumed that in the specimen exists only one defect and the type of the defect is known. The length of such a defect with main orientation in the direction of movement of the solid bar was successfully determined in all cases, whereas the reconstruction in the plane parallel to the conductor upper surface was satisfactory as well. The results have indicated that LET and LFE have a great potential for investigating defects in non-ferromagnetic conductors, especially in laminated composites.

The Lorentz force evaluation (LFE) can be regarded as an interesting alternative to the well-established eddy current testing technique, when in a non-ferromagnetic moving conductor deep internal defects/anomalies have to be identified. In such cases, complicated structures with moving components lead to very high computational costs, i.e., is highly preferable to apply fast numerical models to solve the forward problem. Solving the inverse problem, i.e., the identification of defects or conductivity anomalies, requires efficient solution strategies. In most cases, the number of optimization methods that can be applied for this purpose is limited because usually only derivative-free methods can be used. There is a need of much more research, in particular in the use of stochastic optimization methods for the development of efficient solution strategies. Nevertheless, the application of different inversion algorithms (including improved forward modeling in the defect region) and regularization techniques or usage of more complex permanent magnet configurations (to generate a maximum magnetic flux density) defines the next steps to improve the reconstruction accuracy of the LET and LFE as options of the motion-induced eddy current testing and evaluation.

---

## Applications

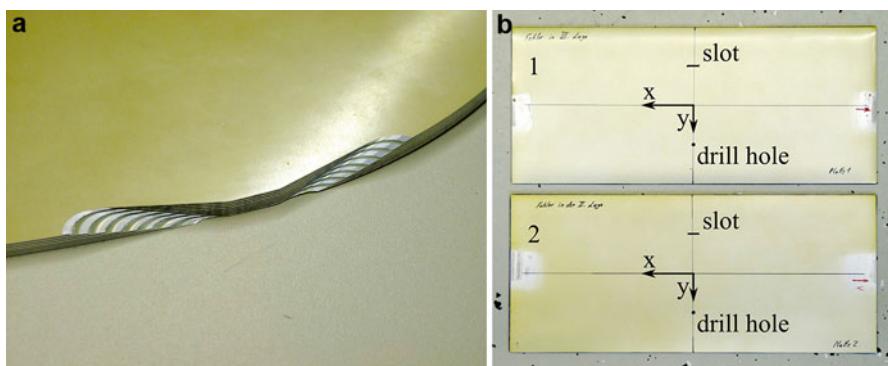
In industry, the market share of composites is increasing fast, and composites are now widely used in the aerospace industry, the automotive industry, the marine industry, etc. They can be found in ship hulls, floor panels, architectural claddings, etc. The composites industry continues to evolve. There is a huge potential for a similar technology shift in the architectural and building and construction segments as the industry takes advantage of the design flexibility, durability, low weight, corrosion resistance, and other properties that composites offer. Composite materials have fueled the growth of new applications in markets such as transportation, construction, corrosion resistance, marine, infrastructure, consumer products, electrics, aerospace, appliances, and business equipment. They are used in industrial applications where corrosion resistance and performance are critical. Fiberglass is often used as the reinforcing fiber. Industrial composite products include underground

storage tanks, scrubbers, piping, fume hoods, water treatment components, and pressure vessels. The aerospace industry, including military and commercial aircraft, is the major customer for advanced composites

Due to the complex structures of the composites with very different and often not precisely known material properties, it is difficult to provide appropriate techniques for testing and evaluation. If the MIECT technique should be applied, only electrically conducting materials can be evaluated, i.e., with NDT methods like LET, the conductive components of the composite devices can be evaluated because there the eddy current flow is induced. Furthermore, the evaluation of composite materials and in particular the identification of defects or conductivity anomalies need the numerical modeling of the eddy current flow in these conducting components. Thus, in the following section, laminated composite materials will be considered.

### Inspection of Glass Laminate Aluminum Reinforced Epoxy

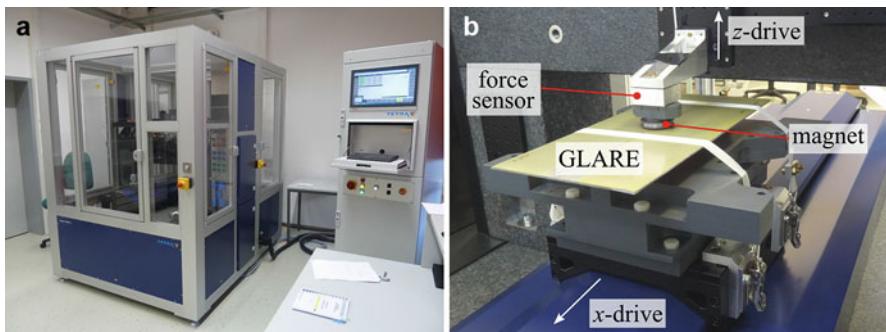
Composite materials can withstand stronger mechanical stress, but at the same time, they are a big challenge for nondestructive testing and evaluation. In this section, measurement results using LET for Glass Laminate Aluminum Reinforced Epoxy (GLARE) samples are presented and compared with simulations (Brauer et al. 2017; Gorges et al. 2016). GLARE is a fiber metal laminate (FML) composed of several very thin layers of metal (usually aluminum) interspersed with layers of glass-fiber bonded together with a matrix such as epoxy. The unidirectional prepreg layers may be aligned in different directions to suit the predicted stress conditions. Figure 15a shows a composition of eight aluminum (light gray) and seven glass-fiber laminate layers (dark gray). In the study presented here, two GLARE samples (aluminum alloy No 3.1354,  $\sigma_0 = 17\text{ MS/m}$ ) shown in Fig. 15b are used. Both specimens are 350 mm × 150 mm and made of 5 aluminum sheets, each 0.4 mm thick, with



**Fig. 15** Test specimen (GLARE): (a) test piece of A380 fuselage (cabin window); (b) two specimens (each 350 mm × 150 mm) with machined defects included in the second aluminum layer (2) and third aluminum layer (1); the same defects in both Al layers: a drill hole (diameter = 2 mm) and a through slot (1 mm × 10 mm)

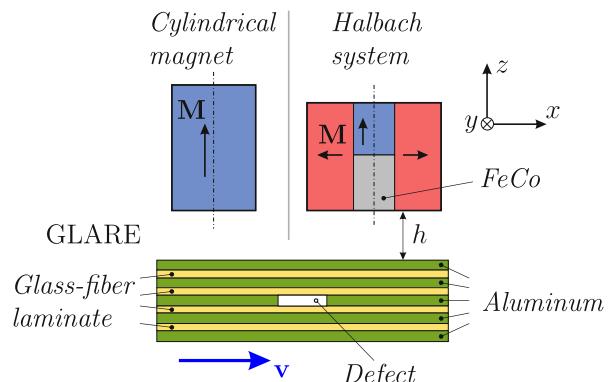
0.25 mm glass-fiber laminate between the sheets, resulting in a total thickness of the specimen of 3 mm. Both specimens have as machined defects a slot ( $10 \text{ mm} \times 1 \text{ mm}$ ) and a drilled hole (diameter = 2 mm) at the marked spots in one of their aluminum layers. Specimen 1 has the defects in the third aluminum layer (defect depth = 1.3 mm) and specimen 2 in the second layer (defect depth = 0.65 mm), counted from the top layer.

The experiments have been performed on the multipurpose measuring platform BASALT-C MMP-15 available at our university laboratory (Fig. 16) (TETRA Gesellschaft 2015). In the study two magnet systems have been used, either a cylindrical permanent magnet or the Halbach system positioned above the GLARE specimen at a lift-off  $h$  (Fig. 17) (Gorges et al. 2016). The sample specimen is mounted onto the  $x$ -slide. The slide moves the specimen with a constant velocity crossing the magnetic field  $B_0$  produced by the magnet system. Due to the relative motion between  $B_0$  and conductive parts, eddy currents are induced in the specimen resulting in the Lorentz force acting on the specimen. The corresponding counterforce  $\mathbf{F} = [F_x, F_y, F_z]^T$  is exerted on the magnet and is measured with a three-axial strain gauge force sensor. The force sensor used during measurements is



**Fig. 16** Measurement setup for GLARE specimen: (a) multipurpose measurement platform MMP-15 and (b) LET measurement of GLARE specimen

**Fig. 17** LET experiment setup for GLARE samples and different magnet systems

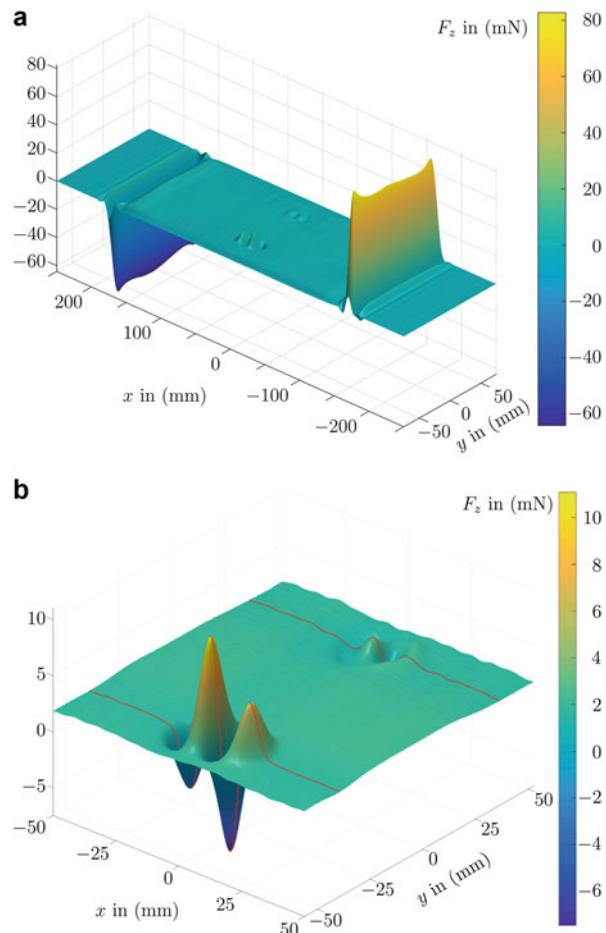


the K3D40, a triaxial sensor with  $\pm 2$  N nominal force in all three directions (ME-Meßsysteme 2014). The LET experiments were carried out at the specimen velocity of  $v = 200$  mm/s and the lift-off distance  $h = 1$  mm. The  $y$ -coordinate was incremented by 0.5 mm over the 150 mm width of the specimen. The cylindrical magnet and the Halbach structure were used to obtain scans of force profiles for various GLARE specimens (Fig. 18). The surface scan leads to a plot of the lift force profile shown in Fig. 18 obtained for the GLARE sample presented in Fig. 15b.

Figure 19 shows the measured drag and lift forces along the centerline of the specimen, where no defect is present. The leading edges of the specimen entered the magnet field at  $x \approx -175$  mm. The trailing edge left the magnet field at  $x \approx 175$  mm. The lift force  $F_z$  has a positive and negative peak at these positions,

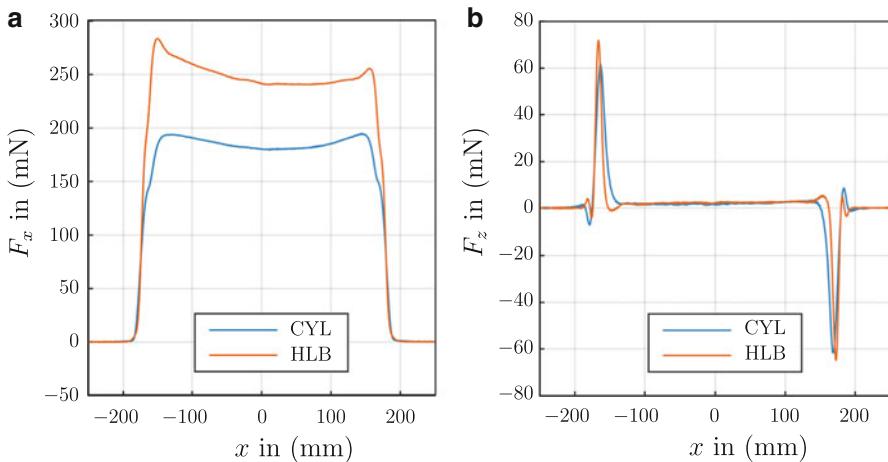
whereas the drag force  $F_x$  rises to a plateau and drops back to zero. Figure 19a shows that the plateau has a dent, which is caused by the uneven surface resulting in a change of the lift-off distance. Both the cylinder magnet and the Halbach system

**Fig. 18** Sample LET scan of lift force profile ( $F_z$ ) for the GLARE specimen with two defects: (a) total scan and (b) zoom into the region around the defects

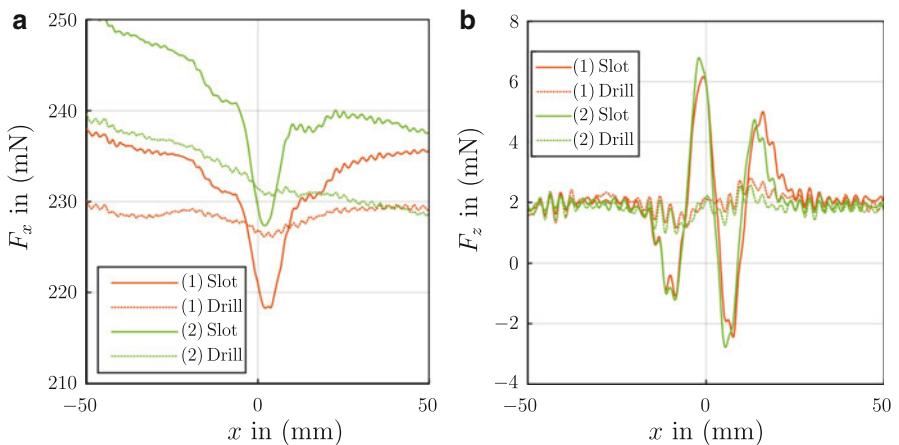


show this behavior. Using the Halbach configuration, the maximum forces are clearly higher, and the general force profile is sharper comparing to the cylindrical magnet. This is the result of the focused magnetic flux density produced by the Halbach system. Because the design of this sensor leads to a field-focusing effect just below the magnet bottom, much higher magnetic flux density values can be achieved. This results in significantly higher Lorentz forces (especially for the drag force) which are about 25%–30% higher than those for the cylindrical permanent magnet.

Figure 20 shows a comparison of the Lorentz force acting on the magnet depending on the defect depths. Here only the results for the Halbach system are shown.



**Fig. 19** Measured forces exerted on the cylindrical magnet (CYL) and Halbach system (HLB)



**Fig. 20** Measurements of the Lorentz force exerted on the Halbach structure by scanning both GLARE samples (see Fig. 15b) which is moving in the  $x$ -direction

The plots are cropped to the range  $-50 \text{ mm} \leq x \leq 50 \text{ mm}$ . The solid lines correspond to the slot defect while the dotted lines to the drill hole defect. The results for GLARE specimen (1) and (2) are depicted with red and green color, respectively.

It can be noted that the deformed uneven surface affects  $F_x$  a lot and makes it difficult to find out, where the deeper defect is located (Fig. 20a). In Fig. 20b, it can be observed that for specimen (2), the deflections caused by the slot defect are a little bit larger than for specimen (1). Since specimen (2) has the defects in the second aluminum layer, they are 0.65 mm closer to the magnet system than in specimen (1), and this indeed results in a stronger deflection in the force. The drill hole is hardly distinguishable in the  $z$ -component. From the drill hole signals, it cannot be concluded which is the deeper defect, since the noise oscillations in the signal are too high (Brauer et al. 2017; Gorges et al. 2016).

It has been shown that the cylindrical Halbach structure has, compared to a cylindrical permanent magnet, a superior performance in detecting small subsurface defects (Weise et al. 2015a). The normalized root-mean-square error (NRMSE) between the numerically calculated and measured Lorentz force signals is used to define the goal function. A finite element model is used to calculate the magnetic flux density, including the extended area approach for the force calculation.

The goal function scanning (GFS) method has been applied to solve the inverse problem described in Fig. 15b, i.e., to identify the two defects at different depths (Mengelkamp 2016; Mengelkamp et al. 2016; Storn and Price 1997). Therefore, the landscape of the goal function is investigated for a set of grid points uniformly distributed in the search space. Since the GFS method is limited to one defect, it is evident to use the geometry parameters of the defect as design variables. Then, the location of the defect, i.e., the  $x$ - and  $y$ -coordinates of its center of gravity, can be determined easily. The defect location is assigned to the position at which the  $\Delta F_x$ -component of the DRS has the largest absolute amplitude.

The GFS is applied to each aluminum layer separately. The results show that there exists one local minimum in each single aluminum layer. Thus, if the optimization function is considered separately for each layer, it is convex. The minimum NRMSE in the five layers corresponds to the correct defective second layer. It has been observed that with the increasing depth of the metal layer, the defect extensions corresponding to the local minima in the single-layer increase. Furthermore, because the goal function is relatively flat, the minimums cannot be distinguished precisely (Mengelkamp 2016). Much more research is required to adapt appropriate solution strategies to this kind of inverse problems.

## Defectoscopy of Friction Stir Welding

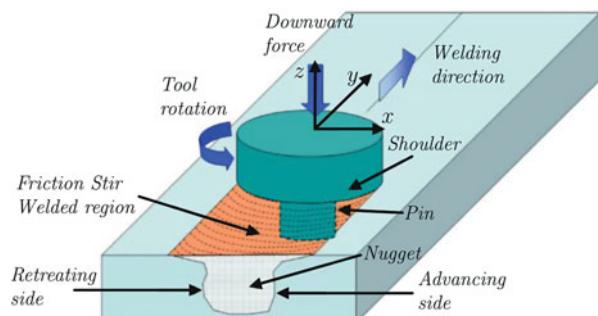
### Friction Stir Welding

Friction stir welding (FSW) is a solid-state joining process that uses a non-consumable tool to join two facing workpieces without melting the workpiece

material. It was invented and experimentally proven at The Welding Institute (TWI), Cambridge/UK, in 1991 (Dawes and Thomas 1996; Thomas et al. 1991). A non-consumable rotating tool with a specially designed pin and shoulder is inserted into the abutting edges of sheets or plates to be joined and traversed along the line of joint (Fig. 21). The tool serves two primary functions: (a) heating of the workpiece and (b) moving the material to produce the joint. The heating is accomplished by friction between the tool and the workpiece and plastic deformation of the workpiece. The localized heating softens the material around the pin, and the combination of tool rotation and translation leads to movement of material from the front of the pin to the back of the pin. As a result of this process, a joint is produced in solid state. During the FSW process, the material undergoes intense plastic deformation at elevated temperatures, resulting in generation of fine and regular recrystallized grains. The fine microstructure in friction stir welds produces good mechanical properties (Mishra and Ma 2005; Mishra et al. 2014).

FSW is emerging as a very effective solid-state joining technique. In a relatively short duration after its invention, several successful applications of FSW have been demonstrated. FSW is primarily used on wrought or extruded aluminum and for structures which need very high weld strength. Therefore, FSW is particularly found in modern shipbuilding, trains, and aerospace applications. In addition to aluminum alloys, friction stir welding has been used successfully to join other metallic materials, such as copper, titanium, steel, magnesium, and composites. Because of high melting point and/or low ductility, successful joining of high melting temperature materials by means of FSW was usually limited to a narrow range of FSW parameters. Preheating is beneficial for improving the weld quality as well as increase in the traverse rate for high melting materials such as steel (Reddy and Reddy 2016). Despite considerable interests in the FSW technology in the past decade, the basic physical understanding of the process is adverse. Some important aspects, including material flow, tool geometry design, wear of welding tool, microstructural stability, and welding of dissimilar alloys and metals, still require deeper understanding. However, it can be observed that new technologies are often commercialized before a fundamental science emphasizing the underlying physics can be developed. This seems to be the case for the FSW technology as well.

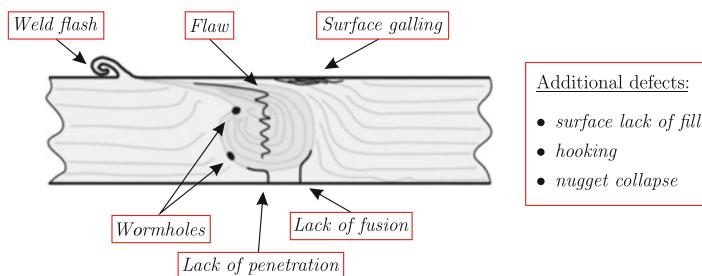
**Fig. 21** Schematic drawing of the principle of friction stir welding (Mishra and Ma 2005)



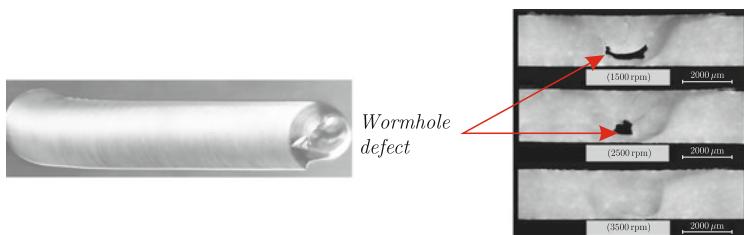
## Imperfections/Defects Caused by Friction Stir Welding

In the different zones of the stirred region, the physical behavior of the material as well as the selected welding parameters during the welding process will cause different potential defects (Fig. 22). The tool geometry is a very important factor in producing sound welds. However, at the present stage, tool designs are generally proprietary to individual researchers, and only limited information is available in the open literature. From the literature it is known that a cylindrical threaded pin and concave shoulder are widely used welding tool features. The welding parameters, including tool rotation rate, traverse speed, spindle tilt angle, and target depth, are crucial to produce sound and defect-free welds. Thus, the design of the tool shoulder together with the applied force pushing the tool on the specimen surface, the rotational speed of the tool, and the speed of welding along the specimen has an important impact on the degree of changes of the material plasticity and on the potential appearance of defects. Thus, FSW can be associated with a number of imperfections or defects if it is not performed properly (Fig. 22).

Insufficient weld temperature, due to low rotational speed or high traverse speed, for example, means that the weld material is unable to accommodate the extensive deformation during welding. This may result in long, tunnel-like defects (so-called wormholes) running along the weld, which usually occur inside the welding seam (Fig. 23). The wormholes appear if process parameters are not correctly chosen and the heat input is insufficient. This happens if either the force the tool is pushing on



**Fig. 22** Typical defects/imperfections which can be found in or around the welding zone (Voellner 2010)



**Fig. 23** Wormholes appearing in the welding zone if the force on the tool or its rotational speed is too low (Voellner 2010)

the material is too low or the rotational speed of the tool is too low. The appearance of wormhole defects can be reduced by increasing the rotational speed. On the other hand, if the generation of wormhole channels is a desired effect, this can be achieved by means of tuning the welding speed. In this way, e.g., cooling channels might be realized inside the welding seam.

### Nondestructive Testing of Friction Stir Welds

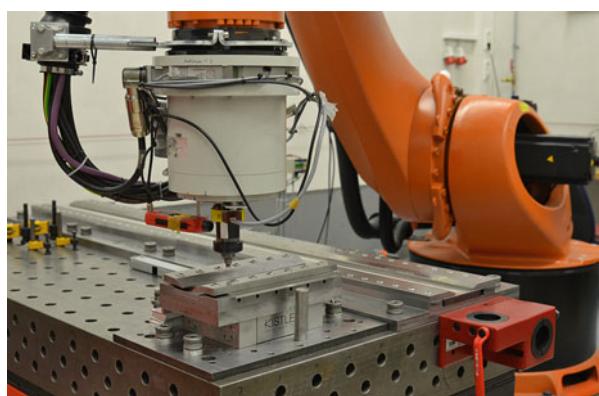
Numerous imperfections or defects have been defined in the literature (Kim et al. 2006; Pietras and Weglowski 2014; Voellner 2010). Unfortunately, these defects cannot be identified in a nondestructive way so far. The following Al-specimens have been welded using the robot-based setup of the Production Engineering Lab at the Technische Universität Ilmenau. This FSW station uses a KUKA six-axis joint-arm robot. Two pieces of Al alloy (thickness = 8 mm) were joint on this FSW system (Fig. 24). The following process parameters have been chosen:

- Material: *Al*-alloy, EN AW 6060 T66
- Dimension: 300 mm × 100 mm × 8 mm
- Feed rate: 400 mm/min
- Tool rotation speed: 2000–3000 rpm
- Tool force: 8 kN
- Surface treatment: milled

The main goal of this study was to analyze whether the MIECT technique is able to evaluate the welding zone properly. It should prove which imperfections/defects can be found in or close to the welding seam, because FSW is well known for its reproducibility and freedom from traditional fusion welding imperfections such as shrinkage cavities or slag inclusions.

If FSW should be widely accepted as a joining method, reliable, but also cost-effective process, specific quality assurance activities have to be developed. So far there is neither a common standard defect catalogue for FSW, which summarizes all relevant irregularities and describes their allowable sizes for different applications,

**Fig. 24** Robot-based FSW of aluminum specimen in the Production Engineering Lab at the Technische Universität Ilmenau, Germany



nor a standardized test specification for FSW welds. Moreover, it is not even fully understood how different imperfections of the weld are affecting its mechanical properties during static and dynamic loads.

Off-line methods include nondestructive tests, commonly applied in welding engineering and in testing the quality of FSW joints. Most popular NDT methods include visual testing, penetrant inspection, ultrasonic examination, X-ray testing, and eddy current testing. Other useful techniques include modern methods such as synchrotron radiation and computer tomography (Pietras and Wegłowski 2014). Standards (EN ISO 25239-5 and AWS D17.3) recommend the following NDT methods for testing FSW welding imperfections (American Welding Society 2016; German Institute for Standardization 2015):

- Visual testing (VT)
- Liquid penetrant inspection (LPI)
- X-ray testing (CT)
- Ultrasonic testing (UT)

In addition, the standard AWS D17.3 enables application of other tests such as acoustic emission, eddy current testing, neutron radiography, leak tests, etc. Individual types of tests have various intended uses and various levels of detectability of specific welding imperfections. Eddy current testing of FSW joints is used, but it is not one of the techniques that are explicitly recommended by the standard for the inspection of FSW joints (German Institute for Standardization 2015). In the case of thin joints, imperfections could be expected in the whole cross section. Eddy current testing of FSW joints requires individually designed transducers, which optimizes the technique (dos Santos et al. 2008; Pitkänen et al. 2014; Rosado et al. 2010). The detection of conventional flaws, like the lack of penetration and lack of fusion, is an important and challenging NDT task. The usage of conventional EC probes demonstrates there is no distinctive signal feature that can allow the distinction between each defect conditions. In order to improve the reliability of nondestructive inspection using eddy current testing in FSW, alternative methods or new probes have to be introduced.

### MIECT Measurements of Friction Stir Welds

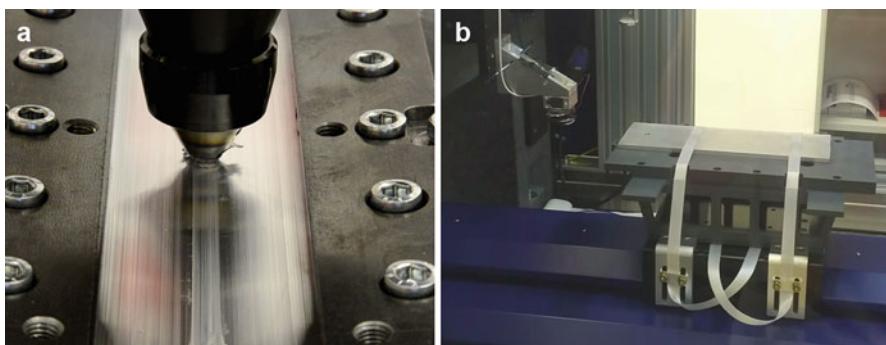
The current situation for nondestructive testing of friction stir welds is characterized by the application of only a few conventional techniques (visual inspection, ultrasound, liquid penetrant inspection, computed tomography), as mentioned in the standards (German Institute for Standardization 2015). Electromagnetic testing methods like eddy current testing are not widely spread. But if metals such as aluminum have to be investigated, it is worth to take into account the utilization of the MIECT technique to identify imperfections or defects in the welding region. It is important to know what happens below the surface and deeper inside the welds or in the welding zones. In the previous section, aluminum alloy specimen has been joint by means of the FSW system available in the Production Engineering Lab. This sample has been used for MIECT measurements on the MMP-15 platform, where

the main goal was to identify different depths of the lack of penetration or perhaps other imperfections/defects in or near the welding zone (Fig. 25). The parameters used for the MIECT measurements are:

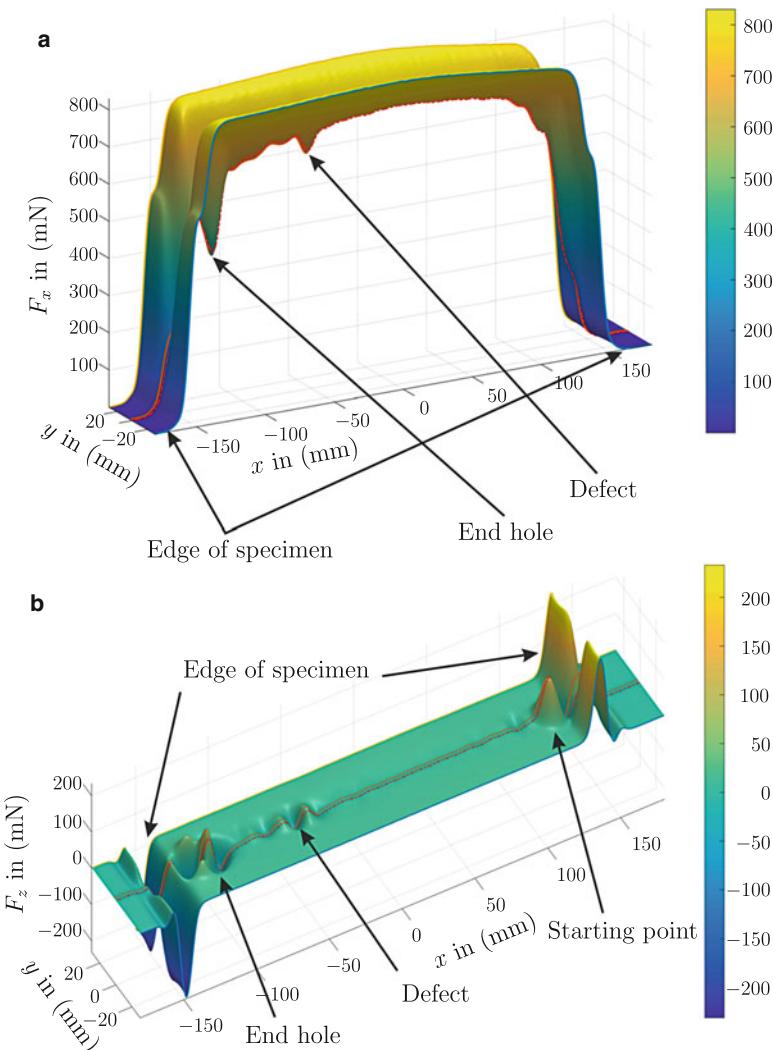
- Magnet: cylindrical Halbach structure (with DiLET coil)
- Orientation of specimen: welded side facing the magnet
- Velocity: 200 mm/s
- Lift-off distance: 1 mm
- Scan width:  $-30 \text{ mm (0.5 mm step)} 30 \text{ mm}$
- Sampling frequency: 1000 Hz
- Low-pass cutoff frequency: 50 Hz
- Scanline repetitions: 25

During the movement of the FSW sample in the vicinity (lift-off distance = 1 mm), the profiles of the Lorentz forces exerted on the permanent magnet have been measured. Both the drag force  $F_x$  and the lift force  $F_z$  were recorded while moving in the  $x$ -direction.

Figure 26 shows the measured force along the length of the specimen, i.e., in a small band (60 mm) along the welding zone around the central line ( $y = 0$ ), just above the welding seam. If there is no disturbance due to imperfections or defects, the drag force profile has a plateau over the entire length of the specimen. In this case, it can be clearly seen that the welding changes the conductivity distribution, compared to the region outside the welding zone. The starting point and the end point of welding can be identified as well as an additional defect close to the end point. It can be observed that at the defect position, the force signal is significantly smaller than in the vicinity. This means that there is a remarkable difference in the electrical conductivity in the welding zone due to friction and stirring or due to defects, e.g., a hole in the seam. This has been confirmed by a follow-up visual inspection.



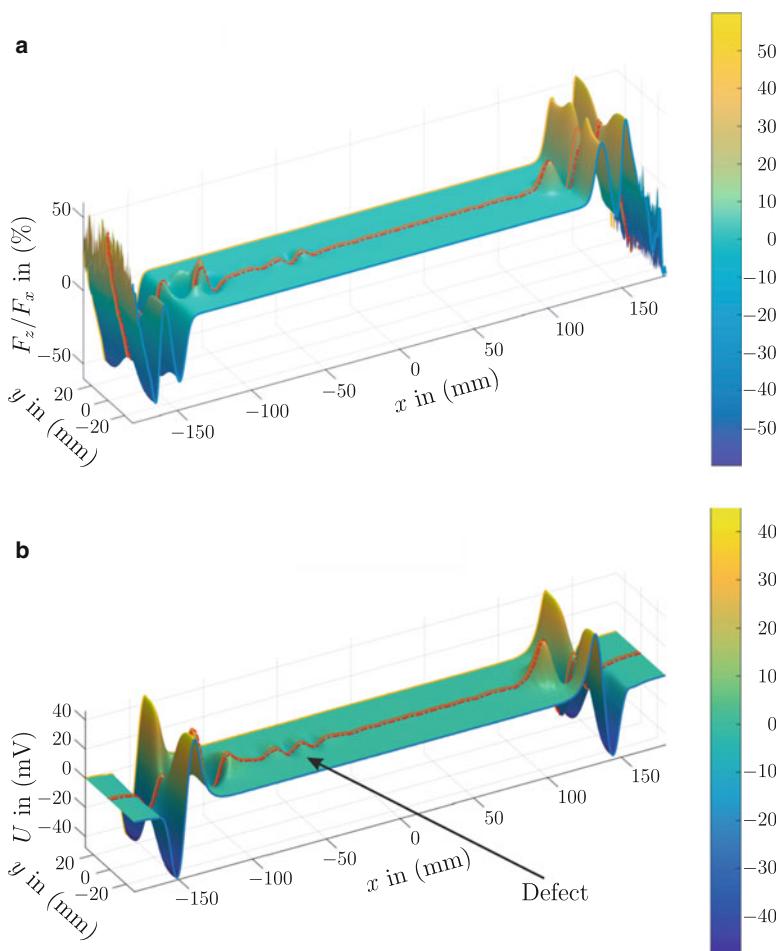
**Fig. 25** (a) Friction stir welding of the *Al* specimen; (b) MIECT measurement of the FSW specimen on MMP-15



**Fig. 26** Force measured along the friction stir weld (red dashed line)

It is obvious that the lift force is more sensitive to the conductivity changes in the welding zone. Again, the edge effect caused by the ends of the sample gives the largest signal. Further, it can be seen that the sample has not been joint to the end of the specimen, because the force signal is zero in those regions. On the other hand, the lift force signals contain obviously more information of the internal defect than the drag force, leading to more details in the force profile around the defect.

The signals shown in Fig. 27 are the result of some post-processing of drag force and lift force. It turned out that the lift-to-drag ratio, i.e., the relation of lift force to drag force, reduces the sensitivity to the lift-off distance and conductivity deviations.



**Fig. 27** Lift-to-drag-force ratio and DiLET signal measured along friction stir weld

This results in data which show higher contrasts than the signals separately. The position of the defect can be identified in this figure at about 2.5 mm beside the centerline (red dashed lines). Furthermore, another near-surface defect has been located close to the starting point. This unexpected finding has been identified afterward as a micro-defect, not visible at the surface.

In a second measurement, a DiLET sensor has been used. This is a differential sensor consisting of a permanent magnet configuration combined with a coil containing a large number of windings. In this sensor system, the motion of the specimen in the static magnetic field will induce a voltage in the coil. This voltage is measured and is used for the defect detection. If the magnet system has been optimized with respect to the desired kind of measurement task, a much higher magnetic flux density can be achieved in the region of interest. The magnitude of the

measured voltage is proportional to the velocity, i.e., increasing the measurement speed will lead to higher voltage signals, presupposing the sampling rate is high enough.

The DiLET sensor generally yields to a higher sensitivity of the entire sensor system, but it leads also to rather noisy data, which can be observed as ripples in Fig. 27b. On the other hand, the voltage profile confirms the position of the defect at 2.5 mm beside the centerline, which is probably a lack of fusion (Fig. 22). This result confirms basically the finding with the common force sensor (Fig. 26).

## Potential Applications of MIECT

### Nondestructive Defect Detection

The general aim of nondestructive testing techniques applied to FSW is the detection of imperfections and/or defects. This includes the determination of the electrical properties (e.g., electrical conductivity) as an alternative for the peak-to-peak measurement, the estimation of the oxide band (so-called kissing bonds), the identification of nugget collapses, or the investigation at both the advancing and the retreating sides of the welding joint. Further, it would be very helpful to provide a technique to identify any lack of penetration from the welding surface in a nondestructive way, especially for thick material components, i.e., much thicker than the penetration depth of the electromagnetic field.

### Material Science

There is a high potential for the development of the MIECT technique if the structure transformation in different materials during welding can be evaluated. The investigation of grain size-affected properties in similar and dissimilar aluminum alloys is a promising field of application, with a high impact on lightweight constructions or aircraft engineering. Furthermore, the interaction between process parameters and grain structure is highly interesting. This include the improvement of the understanding of the necessary structural adjustment for components with high static and dynamic stresses, the entire friction stirring process, or the electrical characterization of the interfaces between different grain sizes.

### Process Control and Monitoring

An important improvement would be the integration of the MIECT into the welding process itself, either for surface-based inspection or for weld root-based inspection. This would generally qualify the whole process monitoring, both with a tracking system for the process control and with an independent, nondestructive weld seam inspection for the quality control of the welding process.

### Ferromagnetic Materials

This chapter focuses on the application of MIECT methods and in particular LET in the framework of non-ferromagnetic materials. However, the moderate rate of change of the magnetic field with respect to the specimen together with the application of NdFeB magnets with high energy densities leads to considerable

advantages in terms of penetration depth, which is the physical limit of traditional ECT. Therefore, the goal is to provide an outlook and to reveal the potential of the MIECT/LET method regarding its usage in combination with ferromagnetic materials. Typical areas of application are, for example, railway testing, steel casting, or pipeline inspection, where safe operation is paramount. In most application scenarios, motion is inherently present and already part of the operation, which privileges MIECT in general especially when it comes to online and in process inspection. Because of distinct differences in the underlying physics, it cannot make claims to completeness since it would require a comprehensive retreatment of the content presented this far. The principled approach of MIECT in case of ferromagnetic specimen resembles the principle of magnetic flux leakage (MFL) (Shi et al. 2015), which is one of the most popular methods of pipeline inspection. In MFL, the specimen is subject to an external magnetic field. Defects are detected on the basis of the leakage of the magnetic field, i.e.,  $\mathbf{B}^{(p,s)}$ , which is sensed by, for example, Hall effect sensors or GMRs (Shi et al. 2015). However, MIECT extends the principle to moving parts.

The nonlinear relationship between the magnetic field and the magnetic flux requires a reformulation of the governing equations. The primary magnetic field  $\mathbf{B}^{(p)}$  is now altered in the presence of the ferromagnetic specimen. Therefore, it can be further divided into a part containing only the magnetic flux generated from the permanent magnet itself as in the case of non-ferromagnetic specimen  $\mathbf{B}^{(p,m)}$  and into a distorted part, resulting from the ferromagnetic specimen  $\mathbf{B}^{(p,s)}$  such that  $\mathbf{B}^{(p)} = \mathbf{B}^{(p,m)} + \mathbf{B}^{(p,s)}$ . The backreaction of the motional eddy currents, i.e., their magnetic flux, is the secondary part  $\mathbf{B}^{(s)}$ . One fundamental difference compared to non-ferromagnetic specimen is that the distorted part of the primary field  $\mathbf{B}^{(p,s)}$  leads to considerable Kelvin forces attracting the magnet to the specimen. In this way, the Kelvin force acts against the Lorentz force and a complex interaction between both forces can be observed. Another elementary difference arises when considering the presence of a defect and the time derivatives of the mentioned field components. Now, in addition to a varying secondary magnetic field from the induced eddy currents ( $\partial\mathbf{B}^{(s)}/\partial t \neq 0$ ), the distorted part of the primary magnetic field changes over time as well ( $\partial\mathbf{B}^{(p,s)}/\partial t \neq 0$ ). The latter plays a key role and privileges the application of differential magnetic field sensors, e.g., coils, to sense field variations in addition to the force as it is presented in section “[Differential Lorentz Force Eddy Current Testing Sensor](#).”

This requires changes of the numerical simulation environments and additional effort to propose efficient approaches since nonlinear magnetic materials involve higher computational cost because of the usage of iterative solvers. Note that some of the semi-analytical approaches include the possibility to model linear ferromagnetic materials (Reitz 1970; Reitz and Davis 1972; Saslow 1992; Ziolkowski 2015). This is incorporated in the analysis of oscillatory motion as well (Amati et al. 2007; Ooi 1977; Ooi and Jain 1979; Weise 2016; Weise et al. 2015b), which could serve as reference solutions for first numerical simulations and implementations thereof.

The availability of efficient numerical approaches would enable the revision of optimal magnet designs by replacing the proposed forward model and a possible

redefinition of the quantity of interest depending on the applied sensor system. The remaining procedure of the proposed optimization strategy remains unaltered because of its general form. In the same course, the uncertainty and sensitivity analysis should be retreated in order to identify the most influencing parameters. The uncertainty analysis of nonlinear magnetic materials is more intricate, but information about its treatment can be found in (Roemer et al. 2014). Note that because of the attraction and magnification of the magnetic flux into the specimen, the useful signal will be influenced positively by being more sensitive with respect to the material properties of the specimen.

The major differences also necessitate a revision of the experimental setup and the applied sensors. Because strong Kelvin forces are present, the lift-off distance between the magnet and the specimen plays a major role during the design process of new systems. In addition to the Lorentz force, magnetic field sensors such as coils, Hall sensors, or GMRs are of particular interest depending on the final application. Perhaps the greatest influence of introducing ferromagnetic materials can be observed in terms of defect evaluation and in solving the inverse problem. The difficulty, as pointed out in the preceding discussion, lies in the estimation of the primary magnetic field distribution  $\mathbf{B}^{(p)} = \mathbf{B}^{(p,m)} + \mathbf{B}^{(p,s)}$  and the associated eddy current distribution. Efficient forward models would again provide an indispensable starting point. A promising solution approach lies in the use of magnetic dipoles in addition to the presently used current dipoles to determine the total magnetic flux and the resulting eddy current distribution.

Notwithstanding the above, the introduction of ferromagnetic materials engenders a series of new challenges. The calculation of the electromagnetic fields is more complex and costly and so are all related tasks depending on those. However, applications such as in-service high-speed railway inspection give reason enough to pursue the effort to further elaborate the MIECT method in this direction.

---

## References

- Achenbach JD (2000) Quantitative nondestructive evaluation. *Int J Solids Struct* 37(1–2):13–27
- Amati N, Tonoli A, Canova A, Cavalli F, Padovani M (2007) Dynamic behavior of torsional eddy-current dampers: sensitivity of the design parameters. *IEEE Trans Magn* 43(7):3266–3277
- American Welding Society (2016) Specification of friction stir welding of aluminum alloys for aerospace. AWS D17.3/D17.3M
- Antunes OJ, Bastos JPA, Sadowski N, Razek A, Santandrea L, Bouillaud F, Rapetti F (2006a) Comparison between non-conforming movement methods. *IEEE Trans Magn* 42(4):599–602
- Antunes OJ, Bastos JPA, Sadowski N, Razek A, Santandrea L, Bouillaud F, Rapetti F (2006b) Torque calculation with conforming and non-conforming movement interface. *IEEE Trans Magn* 42(4):983–986
- Biddlecombe CS, Simkin J, Jay AP, Sykulski JK, Lepaul S (1998) Transient Electromagnetic Analysis Coupled to Electric Circuits and Motion. *IEEE Trans Magn* 34(5):3182–3185
- Binns KJ, Lawrenson PJ, Trowbridge CW (1992) The analytical and numerical solution of electric and magnetic fields. Wiley, Chichester
- Bird J, Lipo TA (2009) Modeling the 3-D rotational and translational motion of a Halbach rotor above a split-sheet guideway. *IEEE Trans Magn* 45(9):3233–3242

- Brauer H, Ziolkowski M (2008) Eddy current testing of metallic sheets with defects using force measurements. *Serb J Electr Eng* 5(1):11–20
- Brauer H, Porzig K, Mengelkamp J, Carlstedt M, Ziolkowski M, Toepfer H (2014) Lorentz force eddy current testing: a novel NDE – technique. *CMPPEL* 33(6):1965–1977
- Brauer H, Gorges S, Ziolkowski M (2017) Bewegungsinduzierte Wirbelstromprüfung von Verbundmaterialien. In: Proceedings of the DGZFP- Jahrestagung, Koblenz, Germany, pp 1–8
- Buffa A, Maday Y, Rapetti F (2000) Calculation of eddy currents in moving structures by a sliding mesh-finite element method. *IEEE Trans Magn* 36(4):1356–1359
- Carlstedt M (2016) A contribution to the experimental validation in Lorentz force eddy current testing. Dissertation, Technische Universität Ilmenau, Ilmenau
- Carlstedt M, Porzig K, Ziolkowski M, Uhlig RP, Brauer H, Toepfer H (2013) Comparison of Lorentz force eddy current testing and common eddy current testing – measurements and simulations. *Stud Appl Electromag* XVII 39(1):218–225
- Carlstedt M, Porzig K, Uhlig RP, Zec M, Ziolkowski M, Brauer H (2014) Application of Lorentz force eddy current testing and eddy current testing on moving nonmagnetic conductors. *Int J Appl Electromagn Mech* 45(1):519–526
- Chady T, Spychalski I (2017) Eddy current transducer with rotating permanent magnets. In: 22nd International Workshop on Electromagnetic Nondestructive Evaluation (ENDE 2017), Saclay, France, p 2
- Chari MVK, Konrad A, Palmo MA, D'Angelo J (1990) Simulation analysis of magnetic sensor for nondestructive testing by boundary element method. *IEEE Trans Magn* 26(2):877–880
- COMSOL Inc., Burlington. Comsol Multiphysics, 2018
- Davat B, Ren Z, Lajoie-Mazenc M (1985) The movement in field modeling. *IEEE Trans Magn* 21(6):2296–2298
- Dawes CJ, Thomas WM (1996) Friction stir process welds aluminum alloys. *Weld J* 75(3):41–45
- Demenko A (1996) Movement simulation in finite element analysis of electric machine dynamics. *IEEE Trans Magn* 32(3):1553–1556
- Donoso G, Ladera CL, Martín P (2011) Damped fall of magnets inside a conducting pipe. *Am J Phys* 79(2):193–200
- dos Santos TG, Ramos PM, dos Santos Vilaca P (2008) Nondestructive testing of friction stir welding: comparison of planar eddy current probes. In: Proceedings of the 16th IMEKO TC4 Symposium, Florence, Italy, pp 507–512
- Geirinhas Ramos HM, Rocha T, Pasadas D, A. Lopes Ribeiro (2013) Velocity induced eddy currents technique to inspect cracks in moving conducting media. In: IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, The Depot, Minneapolis, pp 931–934
- German Institute for Standardization (2015) Ruehrreibschweißen – Aluminium – Teil 5: Qualitäts- und Prüfungsanforderungen. DIN EN ISO 25239-5
- Golovanov C, Coulomb JL, Marechal Y, Meunier G (1998) 3D mesh connection techniques applied to movement simulation. *IEEE Trans Magn* 34(5):3359–3362
- Gorges S, Brauer H, Ziolkowski M, Carlstedt M, Weise K, Schmidt R, Mengelkamp J (2016) Motion-induced eddy current testing of composite materials. In: Proceedings of the 19th World Conference on Non-destructive Testing (WCNDT), Munich, Germany, Fr.1.F
- Haus H, Melcher JR (1989) Electromagnetic fields and energy. Prentice-Hall, Englewood Cliffs
- Hellier CJ (2013) Handbook of nondestructive evaluation, 2nd edn. McGraw-Hill Education LLC, New York
- Ida N (1995) Numerical modeling for non-destructive evaluation, 1st edn. Chapman & Hall, London
- Ida N, Bastos JPA (1997) Electromagnetics and calculation of fields, 2nd edn. New York, NY, Springer
- Jiles DC (1990) Review of magnetic methods for nondestructive evaluation (Part 2). *NDT Int* 23(2):83–92
- Konstantin Weise (2016) Advanced modeling in Lorentz force eddy current testing. Dissertation, Technische Universität Ilmenau, Ilmenau

- Kim YG, Fujii H, Tsumura T, Komazaki T, Nakata K (2006) Three defect types in friction stir welding of aluminum die casting alloy. *Mater Sci Eng A Struct* 415(1–2):250–254
- Kirpo M, Tympel S, Boeck T, Krasnov D, Thess A (2011) Electromagnetic drag on a magnetic dipole near a translating conducting bar. *J Appl Phys* 109(11):113921
- Lai HC, Rodger D, Leonard PJ (1991) A finite element method for problems with moving parts. In: Proceedings of 8th International Conference on Computation in Electromagnetics (CEM91), London, UK, pp 211–213
- Leonard PJ, Lai HC, Hainsworth G, Rodger D, Eastham JF (1993) Analysis of the performance of tubular pulsed coil induction launchers. *IEEE Trans Magn* 29(1):686–690
- Marechal Y, Meunier G, Coulomb JL, Magnin H (1992) A general purpose tool for restoring inter-element continuity. *IEEE Trans Magn* 28(2):1728–1731
- ME-Meßsysteme (2014) Data sheet – K3D40. ME-Meßsysteme GmbH
- Mengelkamp J (2016) Forward and inverse calculation methods for Lorentz force evaluation applied to laminated composites. Dissertation, Technische Universität Ilmenau, Ilmenau
- Mengelkamp J, Ziolkowski M, Weise K, Carlstedt M, Brauer H (2015) Permanent magnet modeling for Lorentz force evaluation. *IEEE Trans Magn* 51(7):6301211
- Mengelkamp J, Lattner D, Haueisen J, Carlstedt M, Weise K, Brauer H, Ziolkowski M, Eichardt R (2016) Lorentz force evaluation with differential evolution. *IEEE Trans Magn* 52(5):6001310
- Mengelkamp J, Carlstedt M, Weise K, Ziolkowski M, Brauer H, Haueisen J (2017) Current density reconstructions for Lorentz force evaluation. *Res Nondestruct Eval* 28(2):76–100
- Mishra RS, Ma ZY (2005) Friction stir welding and processing. *Mater Sci Eng R Rep* 50(1–2):1–78
- Mishra RS, De PS, Kumar N (2014) Friction stir welding and processing: science and engineering. Springer International Publishing, Cham
- Mook G, Hesse O, Uchanin V (2007) Deep penetrating eddy currents and probes. *Materials Testing*, 49(5):258–264
- Mook G, Michel F, Simonin J (2011) Electromagnetic imaging using probe arrays. *Strojníški vestník* 57(3):227–236
- Muramatsu K, Nakata T, Takahashi N, Fujiwara K (1996) Linear AC steady-state eddy current analysis of high speed conductor using moving coordinate system. *IEEE Trans Magn* 32(3):749–752
- Muramatsu K, Takahashi N, Hashio T, Yamada C, Ogawa M, Kobayashi S, Kuwahara T (1999) 3-D eddy current analysis in moving conductor of permanent magnet type of retarder using moving coordinate system. *IEEE Trans Energy Convers* 14(4):1312–1317
- Ooi B-T (1977) A dynamic circuit theory of the repulsive magnetic levitation system. *IEEE Trans Power Appar Syst* 96(4):1094–1100
- Ooi B-T, Jain OP (1979) Force transients at guideway butt joints in repulsive magnetic levitation system. *IEEE Trans Power Appar Syst PAS-98(1)*:323–330
- Petković B (2013) Assessment of linear inverse problems in magnetocardiography and Lorentz force eddy current testing. Dissertation, Technische Universität Ilmenau, Ilmenau
- Petković B, Haueisen J, Zec M, Uhlig RP, Brauer H, Ziolkowski M (2013) Lorentz force evaluation: a new approximation method for defect reconstruction. *NDT & E Int* 59:57–67
- Pietras A, Wegłowski MS (2014) Imperfections in FSW joints and NDT methods of their detection. *Biul Inst Spawalnictwa w Gliwicach* 58(2):23–32
- Pitkänen J, Haapalainen J, Lippinen A, Sarkimo M (2014) NDT of friction stir welding PLFW1 to PLFW5 (FSWL98, FSWL100, FSWL101, FSWL102, FSWL103) NDT Data Report
- Porzig K, Carlstedt M, Ziolkowski M, Brauer H, Toepper H (2014) Reverse engineering of ECT probes for nondestructive evaluation of moving conductors. *AIP Conf Proc* 1581:1519–1525
- Preston TW, Reece ABJ, Sangha PS (1988) Induction motor analysis by time-stepping techniques. *IEEE Trans Magn* 24(1):471–474
- Ramos HG, Lopes Ribeiro A (2014) Present and future impact of magnetic sensors in NDE. *Procedia Eng* 86(1):406–419
- Ramos HG, Rocha T, Pasadas D, Ribeiro AL (2013) Faraday induction effect applied to the detection of defects in a moving plate. *Rev Prog Q* 32(1):1490–1497
- Reddy NR, Reddy GM (2016) Friction stir welding of aluminium alloys – a review. *Int J Mech Eng Technol* 7(2):73–80

- Reitz JR (1970) Forces on moving magnets due to eddy currents. *J Appl Phys* 41(5):2067–2071
- Reitz JR, Davis LC (1972) Force on a rectangular coil moving above a conducting slab. *J Appl Phys* 43(4):1547–1553
- Rocha TJ (2017) Velocity induced eddy current testing. Dissertation, Instituto Superior Te'cnico Lisboa, Lisboa
- Rocha TJ, Ramos HG, Lopes Ribeiro A, Pasadas DJ, Angani CS (2015a) Studies to optimize the probe response for velocity induced eddy current testing in aluminium. *Measurement* 67(1):108–115
- Rocha TJ, Ramos HG, Lopes Ribeiro A, Pasadas DJ (2015b) Magnetic sensors assessment in velocity induced eddy current testing. *Sensors Actuators A Phys* 228(1):55–61
- Rodger D, Eastham J (1985) Characteristics of a linear induction tachometer – a 3D moving conductor eddy current problem. *IEEE Trans Magn* 21(6):2412–2415
- Rodger D, Lai HC, Leonard PJ (1990) Coupled elements for problems involving movement (switched reluctance motor). *IEEE Trans Magn* 26(2):548–550
- Rodger D, Leonard PJ, Eastham JF (1991) Modelling electromagnetic rail launchers at speed using 3D finite elements. *IEEE Trans Magn* 27(1):314–317
- Roemer U, Schoeps S, Weiland T (2014) Approximation of moments for the nonlinear magnetoquasistatic problem with material uncertainties. *IEEE Trans Magn* 50(2):417–420
- Rosado LS, Santos TG, Piedade M's, Ramos PM, Vilaça P (2010) Advanced technique for non-destructive testing of friction stir welding of metals. *Measurement* 43(8):1021–1030
- Saslow WM (1992) Maxwell's theory of eddy currents in thin conducting sheets, and applications to electromagnetic shielding and MAGLEV. *Am J Phys* 60(8):693–711
- Shi Y, Zhang C, Li R, Cai M, Jia G (2015) Theory and application of magnetic flux leakage pipeline detection. *Sensors* 15(12):31036–31055
- Storm R, Price K (1997) Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *J Glob Optim* 11(4):341–359
- Tan Y, Wang X, Moreau R (2015) An innovative contactless method for detecting defects in electrical conductors by measuring a change in electromagnetic torque. *Meas Sci Technol* 26:035 602
- TETRA Gesellschaft für Sensorik, Robotik und Automation mbH. Betriebsanleitung: BASALT-C MMP-15, 2015
- Thess A, Votyakov E, Kolesnikov Y (2006) Lorentz force velocimetry. *Phys Rev Lett* 96(16):164501
- Thess A, Votyakov E, Knaepen B, Zikanov O (2007) Theory of the Lorentz force flowmeter. *New J Phys* 9(8):299
- Thomas WM, Nicholas ED, Needham JC, Murch MG, Temple-Smith P, Dawes CJ (1991) Friction stir butt welding. International Patent Application: PCT/GB92/02203
- Trowbridge CW, Sykulski JK (2006) Some key developments in computational electromagnetics and their attribution. *IEEE Trans Magn* 42(4):503–508
- Uhlig RP (2014) An experimental validation of Lorentz force eddy current testing. Universitätsverlag Ilmenau, Ilmenau
- Uhlig RP, Zec M, Ziolkowski M, Brauer H (2011) Lorentz force eddy current testing: validation of numerical results. *Proc Electrotech Inst* 251:135–145
- Uhlig RP, Zec M, Brauer H, Thess A (2012a) Lorentz force eddy current testing: a prototype model. *J Nondestruct Eval* 31(4):357–372
- Uhlig RP, Zec M, Ziolkowski M, Brauer H, Thess A (2012b) Lorentz force sigmometry: a contactless method for electrical conductivity measurements. *J Appl Phys* 111(9):094914
- Voellner G (2010) Rührreibschweißen mit Schwerlast-Industrierobotern. Forschungsberichte IWB. Herbert Utz Verlag, München
- Weise K, Schmidt R, Carlstedt M, Ziolkowski M, Brauer H, Toepfer H (2015a) Optimal magnet design for Lorentz force eddy current testing. *IEEE Trans Magn* 51(9):6201415
- Weise K, Ziolkowski M, Carlstedt M, Brauer H, Toepfer H (2015b) Oscillatory Motion of Permanent Magnets Above a Conducting Slab. *IEEE Trans Magn* 51(10):7209113

- Yamazaki K (1997) Generalization of 3D eddy current analysis for moving conductors due to coordinate systems and gauge conditions. *IEEE Trans Magn* 33(2):1259–1262
- Yamazaki K (1999) 3D eddy current formulation for moving conductors with variable velocity of coordinate system using edge finite elements. *IEEE Trans Magn* 35(3):1594–1597
- Ying P, Jiangjun R, Zhang Y, Yan G (2007) A composite grid method for moving conductor eddy-current problem. *IEEE Trans Magn* 43(7):3259–3265
- Zec M (2013) Theory and numerical modelling of Lorentz force eddy current testing. Dissertation, Technische Universität Ilmenau
- Zec M, Uhlig RP, Ziolkowski M, Brauer H (2013) Finite element analysis of nondestructive testing eddy current problems with moving parts. *IEEE Trans Magn* 49(8):4785–4794
- Zec M, Uhlig RP, Ziolkowski M, Brauer H (2014) Three-dimensional numerical investigations of Lorentz force eddy current testing. *Stud Appl Electromagn* XVI 38(1):83–93
- Zec M, Uhlig RP, Ziolkowski M, Brauer H (2015) Differentieller sensor, Prüfsystem und Verfahren zur Detektion von Anomalien in elektrisch leitfähigen Materialien. Patent EP2893336 A1, Institut Dr. Foerster GmbH & Co. KG
- Ziolkowski M (2015) Modern methods for selected electromagnetic field problems. Wydawnictwo Uczelniane Zachodniopomorskiego Uniwersytetu Technologicznego, Szczecin
- Ziolkowski M, Brauer H (2010) Fast computation technique of forces acting on moving permanent magnet. *IEEE Trans Magn* 46(8):2927–2930



# Low Field Methods (GMR, Hall Probes, etc.)

23

Vivek T. Rathod, Portia Banerjee, and Yiming Deng

## Contents

Introduction and Overview .....	828
Low-Field NDE Sensor Technologies .....	830
Induction Coils .....	830
Hall Sensors .....	834
Fluxgate .....	835
Magnetoresistance (MR) Magnetometers .....	836
Squid .....	838
Eddy Current Magneto-Optic Imaging (MOI) Sensors .....	839
NDE Applications and Beyond .....	839
Pipeline and Tube Defects Inspection .....	840
Subsurface Inspection .....	849
Defects in Permanent Magnets .....	852
Buried Features .....	854
Damages in Wire Ropes .....	855
Crack Detection and Profiling .....	860
Composite and Sandwich Structures .....	861
Multilayer Joints .....	865
Mechanical Properties and Stress .....	867
Rail Wheel Defects .....	870
Metrology .....	873
Clinical Applications .....	873
Summary .....	874
References .....	875

---

V. T. Rathod · Y. Deng (✉)

NDE Laboratory, Department of Electrical and Computer Engineering, Michigan State University,  
East Lansing, MI, USA

e-mail: [dengyimi@egr.msu.edu](mailto:dengyimi@egr.msu.edu)

P. Banerjee

SGT Inc., NASA Ames Research Center, Mountain View, CA, USA

e-mail: [banerj20@egr.msu.edu](mailto:banerj20@egr.msu.edu)

**Abstract**

Low-field nondestructive evaluation techniques have witnessed considerable improvements with the development of highly sensitive magnetic field sensors. Low-field techniques have brought several advantages like high resolution, high signal-to-noise ratio, and improved detection capabilities. This chapter begins with the importance of high-frequency eddy current testing and its limitation due to skin effect. It further presents the recently developed pulsed eddy current-based NDE to overcome the skin effect. Subsequently, this chapter presents the developments on increasing the sensitivity of low-field sensors by developing new probe designs and materials. A subsection presents the principles of operation of the available low-field sensors. The implementation of these low-field sensors in nondestructive evaluation is discussed. It also presents methods to reduce the sensor size for low footprints and increase the sensitivity. The chapter discusses the application of the low-field sensors in a detailed subsection with details on the implementation, its capabilities, and advantages. Finally, the chapter discusses the nondestructive applications of the low-field sensors with details on the implementation, its capabilities, and advantages.

---

**Introduction and Overview**

Periodic inspection of large structures like heat exchangers, pipelines, wind turbines, offshore structures, aircraft, and marine structures requires noncontact inspection and high-speed capabilities (Afzal and Udpal 2002; Xu et al. 2012). Compared to the popular noncontact ultrasound and X-ray, electromagnetic (EM) techniques offer low-cost solutions satisfying these criteria. Eddy current testing (ECT) is popular for its capability to detect subsurface defects in conducting materials (Jander et al. 2005 Liu et al. 2009; Deng and Liu 2011; Garcia-Martin et al. 2011). In addition, it is capable of material characterization and metrology, which are used as testing parameters in nondestructive evaluation (NDE). Paramagnetic materials like aluminum, titanium, and copper can be inspected rapidly, safely with ease of implementation (Le et al. 2013). The state of the art has been established using electromagnetic principles. Modern developments of the probes involve the development of computer models to understand the principles of the technique. Faraday's law and Ampere's law are used to derive the governing equations for electric current sensing and to model new probes.

The changes in the magnetic field around the defect provide information of location, shape, and size. The accuracy and effectiveness of ECT depend on the sensitivity of the receiving sensor to detect or map the magnetic field changes caused by defects. Noting the advantages of magnetoresistive (MR) sensors (Ye et al. 2016), the rapid development of giant magnetoresistance (GMR) sensor has increased the sensitivity many folds. Recent development reports on tunnel magnetoresistance (TMR) sensors have promised a considerable increase in the sensitivity with a drastic reduction in power consumption.

High frequencies in the range of 10–100 kHz are used to increase the voltage in the pickup coil. However, the depth of defect detection diminishes due to the skin effect (Thollon and Burais 1995; Ramos et al. 2009). The skin effect is a

phenomenon whereby the eddy currents induced in the specimen concentrate on the surface and decrease with the depth along the  $z$ -direction exponentially. For the current flowing only in the  $x$ -direction  $\mathbf{J} = J_x \mathbf{u}_x$ , the current density at the surface and its variation along the  $z$ -axis is given by

$$J_x(z,t) = J_0 e^{-z/\delta} \cos\left(\omega t + \alpha_0 - \frac{z}{\delta}\right), \quad (1)$$

where  $J_0$  is the current density at  $z = 0$ ,  $\delta$  is the skin depth, and  $\alpha_0$  is the phase at  $t, z = 0$ . Figure 1 shows the field variation in the aluminum alloy (Al 2024 T3) with a conductivity  $\sigma = 1.75 \text{ MSm}^{-1}$ .

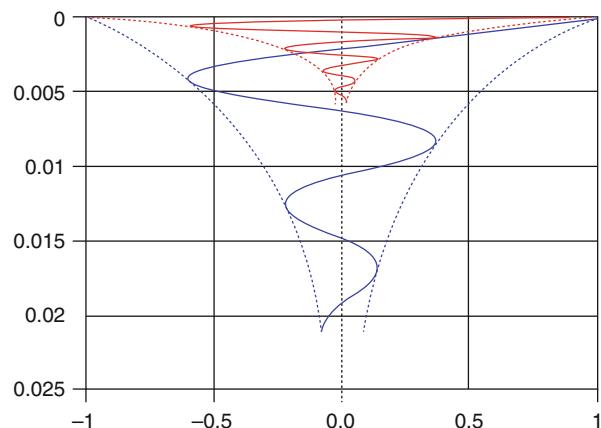
The eddy current density in the specimen below the surface representing the amplitude in the above Eq. (1) falls to  $1/e$  below the surface at a certain depth called the skin depth  $\delta$  given by

$$\delta = \sqrt{\frac{2}{\mu\omega\sigma}}, \quad (2)$$

where  $\omega$  is the angular frequency of the sinusoidal magnetic field.  $\mu$  and  $\sigma$  are the magnetic permeability and the conductivity of the medium, respectively. The induced eddy currents and the associated magnetic field penetrate to a specific depth with a strength dictated by the skin depth. Hence, the inspection depth can be adjusted by proper choice of the operation frequency to detect surface or subsurface cracks. Table 1 shows the numerical values of the skin depth for pure aluminum and aluminum alloy (Al 2024 T3). The numerical values clearly show a drastic reduction of the skin depth at a higher frequency of 10,000 Hz. At 200 Hz, since most of the inner material contributes to the current flow, differences in the magnetic field due to surface cracks will be negligible.

Pulsed eddy current overcomes the limitations due to skin effect by generating the current in the form of pulses or transients instead of time-varying fixed frequency

**Fig. 1** Electromagnetic field variation along the depth ( $z$ -axis) in an aluminum alloy (Al 2024 T3) with a conductivity  $\sigma = 1.75 \text{ MSm}^{-1}$ . (Reproduced from Ramos and Riberio (2014))



**Table 1** Skin depth for aluminum and aluminum alloy at low and high frequency (Ramos et al. 2009)

Material	Skin depth at 200 Hz	Skin depth at 10000 Hz
Pure aluminum	5.990 mm	0.847 mm
Aluminum alloy	8.510 mm	0.120 mm

signal (Smith and Hugo 2001; Sophian et al. 2001). The earlier response in the time domain contains the information of the superficial material, whereas the later response contains information of the deeper parts of the material. The pulse waveform is represented by the Fourier series given by

$$I(t) = I_{SS} \left[ \left( 1 - e^{-t/a} \right) u(t) + \left( e^{-(t-\tau)/a} - 1 \right) u(t-\tau) \right] (0 \leq t \leq T), \quad (3)$$

where  $a$ ,  $I_{SS}$ ,  $\tau$ , and  $T$  are the rising constant, steady-state value, duration time, and period of the excitation current, respectively. Here  $u(t)$  is a unit step function. The broadband, pulsed eddy current response carries the subsurface layer information with increased noise levels for increasing depth. Thus, present developments focus on increasing the sensitivity by developing new probe designs and materials. This chapter summarizes the sensors available to detect the magnetic field and the principles associated with their development in section “[Low-Field NDE Sensor Technologies](#).” Section “[NDE Applications and Beyond](#)” elaborates on the nondestructive evaluation techniques developed to detect common types of defects, to characterize material, and for metrology applications.

---

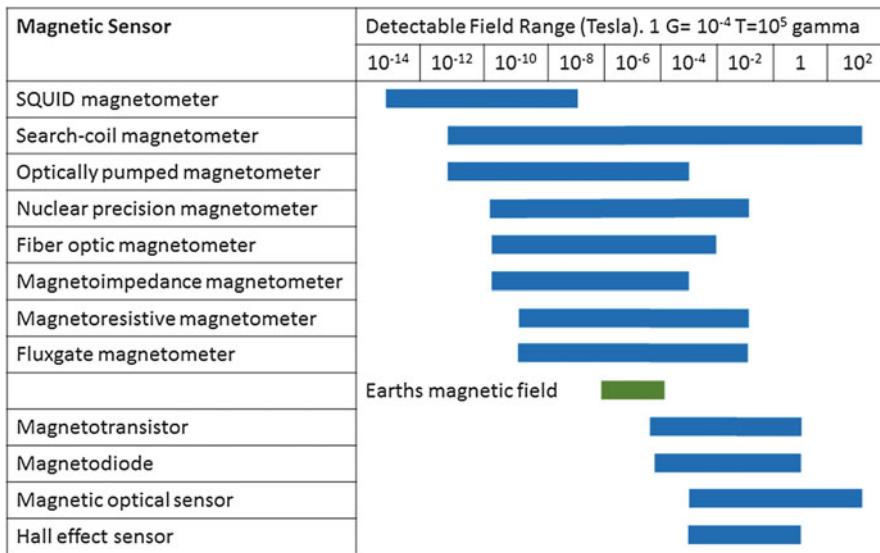
## Low-Field NDE Sensor Technologies

The past 60 years have witnessed several important developments of magnetic sensors like inductive coils, proton precession magnetometers, fluxgate magnetometers, superconducting quantum interference devices (SQUIDs), Hall effect sensors, and magneto-resistors (Kapsalis 2017; Ramos and Ribeiro 2014). The recent trend is to miniaturize them to achieve low footprints and increase the sensitivity. Emphasis is placed on low-field measurement capabilities where a high signal-to-noise ratio (SNR) is desired. Table 2 shows the range of magnetic field sensing of the sensors that are reported or commercially available. Table 3 lists the sensors developed to measure low magnetic fields with their operating principles and specifications.

## Induction Coils

### Introduction

One of the oldest and well-known categories of low-field magnetic sensors is the induction coil, also known as search coil sensors or pickup coil sensors (Campbell 1969). Operating on the principle of Faraday’s law of induction as stated in Eq. (4),

**Table 2** Comparison of magnetic field ranges of magnetic sensor technologies (Caruso et al. 1998; Kaur et al. 2018b)

these sensors, comprising a winding wire either with or without a ferromagnetic core, are the simplest to design with lowest manufacturing costs. It is evident from Eq. (4) that when a conducting coil is near an alternating magnetic field, current is induced in the conducting coil and the generated voltage  $V$  is proportional to the rate of change in the magnetic field. Here,  $\Phi$  denotes the magnetic flux,  $B$  is flux density,  $H$  is magnetic field strength,  $A$  is the area of the coil,  $n$  is the number of turns in the coil winding and  $\mu_0 = 4\pi \times 10^{-7}$  H/m is the permeability of free space.

$$V = -n \frac{d\Phi}{dt} = -nA \frac{dB}{dt} = -\mu_0 n A \frac{dH}{dt}. \quad (4)$$

A conducting coil can generate voltage and/or current in presence of any magnetic field, and therefore induction coil sensors are the most versatile as they can cover almost the entire detectable range of all magnetic field sensors. On the one hand, they can measure very low magnetic field (sensitivity of air-coil  $\sim 0.3$  pT/ $\sqrt{\text{Hz}}$  at 20 Hz, with ferromagnetic core  $\sim 2.5$  pT/ $\sqrt{\text{Hz}}$  at 1 Hz), while on the other hand, they are capable of measuring very high magnetic fields. However, as inferred from Eq. (4), large sensitivity requires a large number of turns ( $n$ ) and a large active area  $A$  of the coil. For example, in geophysical applications, air core sensors with diameters larger than 1 m and weighing close to 100 Kg have been used (Campbell 1969). Therefore, optimization of coil dimensions is an important aspect to note while manufacturing induction coil sensors tailored toward specific applications.

**Table 3** Low-field sensor details

Sensor type	Mode of operation	Measure parameter	Frequency range	Operating field range	Sensitivity	Temperature drift	Resolution	Typical sensor size	Power	Reference or product name
Induction coil	Electromagnetic induction	$\frac{d\phi}{dt}$	—	1 pT–1 mT	—	—	<100 ft Hz <sup>-1/2</sup>	—	—	Campbell (1969)
Hall	Hall effect	<i>B</i>	—	±420 G	5 mV/G	0–0.08%/°C Sensitivity and –0.02–0.08%/°C °C Vnull drift	—	4.06 mm × 1.57 mm × 3 mm	5 V, 7 mA	(SS490) from Honeywell
Fluxgate	Second harmonic principle based on saturation of ferromagnetic core	<i>H</i>	0–12 kHz	±100 μT	100 mV/μT	—	100 pT–10 nT	Dia. 24.5 mm, Length 218 mm	±15 V, 60 mA	(Mag639) from Bartington Instruments
AMR	Anisotropic magnetoresistance	<i>H</i>	58 MHz	±200 μT	0.16 mV/ μT at 5 V	3900 ppm/K Offset and 3000 ppm/K Sensitivity	—	—	—	(H1001/1002) from Honeywell
GMR	Giant magnetoresistance	<i>H</i>	0–1 MHz	1.5–10.5 Oe	3–4.2 mV/ Oe	—	—	1.1 mm × 1.1 mm	100–675 mW	(AA002–02E) from NVE corp
TMR	Tunneling magnetoresistance	<i>H</i>	—	±30 Oe (left) and ±200 Oe (right)	4.9 mV/Oe	—	—	6 mm × 5 mm × 1.5 mm	1 V, 11 μA	(TMR2102) from American Electronic Components Incorporated
SQUID	Josephson junction and flux quantization	∅	—	1 fT–1 nT	1.6 pT Hz <sup>-1/2</sup>	—	—	10.5 mm × 10.5 mm × 10.5 mm	—	Supracon 3Dgreen; Campbell (1969)
MOI	Faraday rotation effect	∅	5–200 kHz	—	—	—	—	—	—	MOI 308/7 from Qi2

Note: 1 Oe (Oersted) = 1 Gauss in air = 0.1 milli-Tesla = 79.8 A/m  
*B*-magnetic flux density,  $\emptyset$ -magnetic flux, *H*-magnetic field intensity

## Working Principle and State-of-the-Art

As mentioned previously, induction coils can be broadly classified into two types: (1) air core and (2) ferromagnetic core. Sensor design and operating principle for both types have been presented and derived in Tumanski (2007). It is known that in the presence of an oscillating magnetic flux density defined by  $B = B_m \sin(\omega t)$ , the sensitivity ( $S$ ) of an air-coil sensor in the form of a ring with diameter  $D$  can be derived as Eq. (5), where  $H$  is the magnetic field strength;  $D$  and  $D_i$  are the outer and inner coil diameter, respectively;  $f$  is the frequency of the measured field; and  $n$  is the number of turns in the coil sensor.

$$S = \frac{V}{H} = \frac{10^{-7} \pi^3 f n (D + D_i)^2}{2}. \quad (5)$$

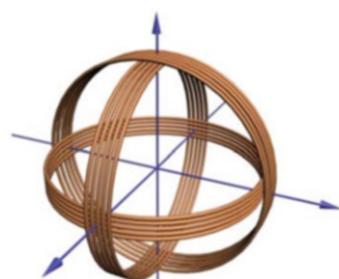
Further, the resolution of an air-coil sensor in terms of signal-to-noise ratio (SNR) can be calculated according to Eq. (6).

$$SNR = \frac{\pi^3 10^{-7}}{8} \frac{f}{\sqrt{\Delta f}} \frac{\sqrt{l(D + D_i)} \sqrt{D^2 - D_i^2} H}{\sqrt{k^2 k_B T \rho}}, \quad (6)$$

where  $k = l(D - D_i)/2nd^2$ ,  $T$  is the temperature (coil gets heated as current flows through it leading to thermal noise and affects the sensing resolution),  $\Delta f$  is the frequency bandwidth of the oscillating magnetic field, and Boltzmann factor  $k_B = 1.38 \times 10^{-23} \text{ WsK}^{-1}$ . Thus, to achieve a higher sensitivity as well as resolution, the diameter of the coil and the number of turns have to be increased. Design of an induction coil magnetometer for measuring stray components of earth's magnetic field in the bandwidth of 0.001–10 Hz with a resolution of 1 pT–1 nT is described in Lenz and Edelstein (2006). Another example of a three-axis search coil magnetometer is presented in Tumanski (2007) capable of measuring magnetic fields between 20 Hz and 20 kHz with a low noise level below 170 dB/100  $\mu\text{T}$  and is shown in Fig. 2.

High-dimensional requirements of air-coil sensors can be managed by inserting high-permeability materials such as ferromagnetic cores inside the coils, which act as flux concentrators. Equation (4) for voltage generation of for induction coil with ferromagnetic core sensors can be modified as

**Fig. 2** An example of three-coil sensor. (Reproduced from Tumanski (2007) with permission from IOP)

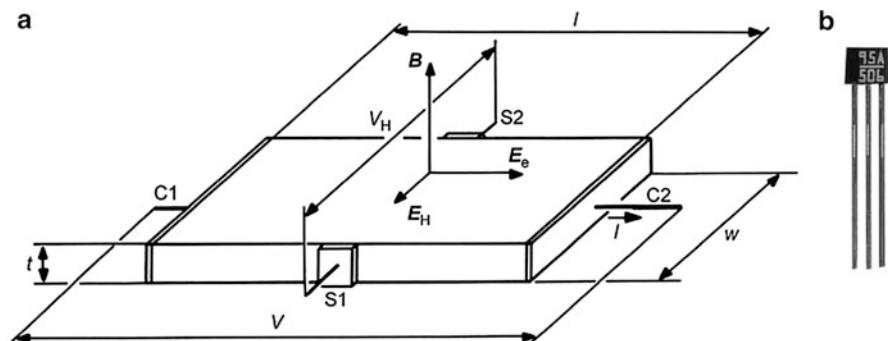


$$V = -\mu_0 \mu_r n A \frac{dH}{dt}, \quad (7)$$

where  $\mu_r$  is the relative permeability of soft magnetic materials (usually larger than  $10^5$ ) which can increase the sensitivity of the coil significantly without the need to increase the area or diameter of the coil. However, while manufacturing ferromagnetic cores for induction coil sensors, it is important to note that the resultant permeability of the core  $\mu_c$  can be much lower than its  $\mu_r$ , owing to the demagnetizing field effect. For example, for an elliptical core, the demagnetizing factor ( $N$ ) is dependent on its length and diameter, and hence usually a long and thinner core is used in these sensors. By using, an amorphous ribbon made of Metglas 2714F with an aspect ratio equal to 30, the sensitivity of the coil could be enhanced 300 times compared to the air-coil sensor (Prance et al. 2000).

## Hall Sensors

Hall sensors are made of semiconductor materials and operate based on the principle of Hall effect. An electric potential difference is produced across the semiconductor when a magnetic field is applied perpendicular to the direction of the current flow. The effect occurs with a direct or alternating field at the same frequency. A typical rectangular plate-type Hall effect sensor is shown in Fig. 3. The dimensions shown can be as small as  $t = 10 \mu\text{m}$ ,  $l = 100 \mu\text{m}$ , and  $w = 200 \mu\text{m}$ . A bias voltage  $V$  is applied across the two contacts C1 and C2 resulting in an electric field  $E_e$  forcing a current  $I$ . In this state, when the plate is exposed to an external perpendicular magnetic field  $B$ , a Hall electric field  $E_H$  results. The Hall electric field generates a voltage  $V_H$  across the sensor contacts S1 and S2 (Popovic 2003) given by



**Fig. 3** (a) Schematic of a rectangular plate-type Hall effect sensor. Modern sensors can have microscopic dimensions like  $t = 10 \mu\text{m}$ ,  $l = 100 \mu\text{m}$ , and  $w = 200 \mu\text{m}$ . (Reproduced from Popovic (2003) with permission from Taylor and Francis). (b) Hall sensor SS495. (Reproduced from Li et al. (2018) with permission from IEEE)

$$V_H \simeq \mu \frac{w}{l} VB \simeq \frac{1}{qnt} IB, \quad (8)$$

where  $q$  denotes the magnitude of electron charge and  $n$  is the carrier concentration of the semiconductor plate. The above formula is not applicable for Hall sensors of different shape and large contacts. The external magnetic flux density of the order of 100 nT with a resolution of 1 nT can be measured with simple acquisition circuits.

## Fluxgate

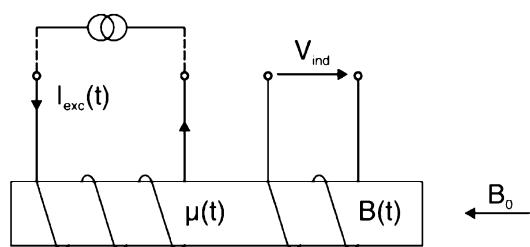
Fluxgate is a vector device capable of measuring DC or low-frequency AC magnetic fields. In conjunction with pulsed eddy currents, the fluxgate magnetometers can measure the decay of the magnetic field  $B_Z$  instead of the derivative  $\partial B_Z / \partial z$  of induction coils. The derivative of the magnetic field decays much faster, due to which the induction coils are less sensitive to deeper layer defects. Thus, fluxgate sensors can provide information of the deeper layers with greater signal-to-noise ratio (SNR). The excitation current  $I_{exc}$  through the excitation coil produces a field that saturates the soft magnetic core material as shown in Fig. 4. The saturation results in the “gating” of the flux and hence the name *fluxgate*. The core permeability drops down decreasing the magnetic field and associated DC flux. The output voltage  $V_i$  is induced in the pickup coil at second and higher harmonics in the presence of the measured field. The general induction sensor equation is given by (Ripka 2003)

$$-V_i = \frac{d\emptyset}{dt} = NA\mu_0\mu_r \frac{dH}{dt} + NA\mu_0 H \frac{d\mu_r}{dt}, \quad (9)$$

where  $\emptyset$  is magnetic flux,  $\mu_r$  is relative permeability,  $\mu_0$  is the permeability in the air,  $N$  is the number of turns,  $A$  is the area of the coil, and  $H$  is the magnetic field strength. The first term appears due to interference. The second term is an important term due to the fluxgate effect of changing core permeability due to periodic change in the excitation field. Equation (9) is applicable for rod-type sensors.

For ring-type sensors, the effective demagnetization factor  $D$  is considered with the field in the core being lower than the field in the air  $H_0$ . In this case, the flux density is  $B = \mu_r \mu_0 H_0 / [1 + D(\mu_r - 1)]$ , which leads to the fluxgate voltage given by (Ripka 1992)

**Fig. 4** The operating principle of the fluxgate sensor. (Reproduced from Ripka (2003) with permission from Elsevier)



$$V_i = NA \frac{dB}{dt} = NA\mu_0 H_0 \frac{1 - D}{[1 + D(\mu_r - 1)]^2} d\mu_r/dt. \quad (10)$$

Clarke (Clarke 1999) has extensively reported the demagnetization factor for various cases. The voltage sensitivity increases with the number of turns and decreases with the demagnetization factor and coil parasitic capacitance. The voltage output is often tuned by a parallel capacitance to minimize distortions leading to the false sensor output and reduced power consumption (Ripka and Primdahl 2000). The presence of defects perturbs the decay of eddy currents, and the fluxgate sensor measures its associated magnetic field. A phase-sensitive detector demodulates the resulting second harmonic voltage at the sensor output to DC or very low-frequency signals. An integrator provides the gain, and an amplifier senses the feedback current across a resistor as the magnetometer's output. Miniaturizing the fluxgate length to fabricate sensor arrays is limited due to the rapidly increasing noise levels. Three single axis orthogonal Helmholtz coils are used to compensate for errors due to the sensitivity to a magnetic field perpendicular to the sensing axis called the *cross-field effect*. Digital magnetometers and pulse-width modulators still have drawbacks like increased noise levels, limited linearity, and higher power consumption. Comparatively, the AMR sensors have better parameters (size < 5 mm), but fluxgate sensors are still popular due to low cost and simplicity. Table 1 details the typical parameters of AMR and fluxgate sensors.

## Magnetoresistance (MR) Magnetometers

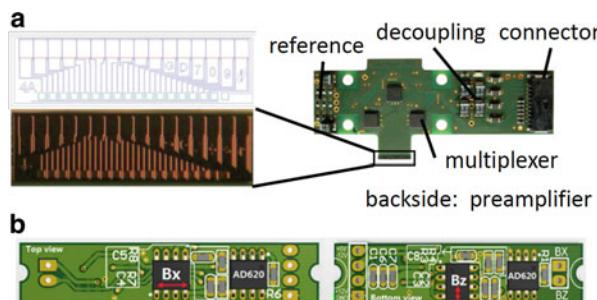
Magnetoresistance sensors have evolved with the developments in solid-state materials and are replacing the traditional inductive sensors with their ease of use, high resolution, and low cost. These devices operate on the principle of magnetoresistance, which is the property of some materials to change their resistance when exposed to a magnetic field. Any conductor exhibits MR effect, but it is negligible. For this reason, the effect in these materials is called ordinary magnetoresistance. Some magnetic materials are anisotropic to magnetic field direction offering a large magnetostrictive effect and can be used in sensor applications. Magnetoresistance is quantified as a ratio defined by

$$MR = (R_{MAX} - R_{MIN})/R_{MIN}, \quad (11)$$

where  $R_{MAX}$  and  $R_{MIN}$  are the maximum and minimum resistances, respectively.

## Anisotropic Magnetoresistance (AMR) Magnetometers

The anisotropic magnetoresistance (AMR) is a change in the scattering cross section of atomic orbitals due to the magnetic field. The scattering causes resistance, which is maximum when the magnetization direction is parallel to the direction of the current. Maximum linearity is achieved when the magnetization is at 45° to the current. Thus, linear AMR sensors are fabricated by patterning highly conductive



**Fig. 5** (a) Left: GMR array with 16 elements to measure the normal field. Right: A board with GMR arrays housing 48 GMR elements and multiplexed read-out electronics. (Reproduced from Pelkner et al. (2014) with permission from AIP). (b) Front and back view of the TMR sensor module. (Reproduced from Yuan et al. (2018) with permission from Elsevier)

diagonal shunts on the AMR material. The directional dependence of change in resistance to magnetic field led to the association of the term “anisotropic.” The application of external magnetic field changes the resistance. The magnetoresistance is small ( $MR = 1\text{--}2\%$ ) compared to the newly developed giant magnetoresistance (GMR) sensors, which is explained next.

### Giant Magnetoresistance (GMR) Magnetometers

Recent advances in nano-thin film technology have helped in the development of GMR sensors. These sensors have very large magnetoresistance of  $MR = 20\text{--}50\%$  exceeding the values possessed by AMR devices. Thereby the term “giant” is associated with the name. The GMR effect is realized by depositing two to three layers of ferromagnetic metal like NiFe and CoFe with a separation of non-magnetic layers made of Cu, Au, or Ru. The nonmagnetic spacer layers are very thin to provide a free path to the spin-polarized electrons to pass to other layers before their polarization is disturbed by scattering (Smith et al. 2003). The layers create magnetic direction, and when they are aligned with the direction of the magnetic field, the electrons pass freely. The high sensitivity of GMR has led to their use in the read heads of hard disk drives. GMR has high bandwidth, high sensitivity, and low power consumption. Figure 5(a) shows an individual GMR sensor developed to detect defects and image them.

### Tunneling Magnetoresistance (TMR) Magnetometers

Tunneling magnetoresistance (TMR) have been developed that have ultrathin insulating layers to separate the magnetic layers from the conductors. Electrons pass through the insulator by quantum mechanical tunneling. The thin layer can easily suffer damage by voltage above a few hundred mV. TMR has a larger magnetoresistance  $MR = 50\text{--}60\%$  compared to GMR’s  $MR = 50\%$ . Figure 5(b) shows the front and back view of a typical printed circuit module with TMR sensors.

## Squid

One of the most powerful NDE technologies that enable detection of extremely small changes in the magnetic field, even lower than the detectability of fluxgate sensors, is the SQUID. The advantages of the SQUID in NDE are its high sensitivity ( $\approx 10\text{--}100 \text{ fT Hz}^{-1/2}$ ), wide bandwidth (from DC to 10 kHz), and a broad dynamic range ( $>80 \text{ dB}$ ). Owing to its high sensitivity and capability to operate at near-zero frequency, SQUIDs allow penetration depths higher than any other low-field electromagnetic methods. Further, it is intrinsically quantitative in nature, which benefits not only anomaly detection but also sizing and probability-of-detection (POD) analysis of flaws.

SQUIDs operate on the principle of superconductivity in the Josephson's junction. Superconducting materials, when cooled down beyond a critical temperature ( $T_c$ ), can allow a DC current (maximum current of  $I_c$ ) to flow through them with zero resistance, i.e., constant current flows without an external power source. Niobium and  $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$  are the predominant superconducting materials used in industrial SQUIDs. Niobium, a low  $T_c$  material, is the most common one with  $T_c = 9.25 \text{ K}$  which is cooled by immersion into liquid helium (having a boiling point of 4.7 K). On the other hand,  $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ , being a high  $T_c$  material, has  $T_c = 93 \text{ K}$  and is cooled using liquid nitrogen whose boiling point is much higher than liquid helium, which is much cheaper and easily available. However, high  $T_c$  SQUIDs may encounter poor magnetic detectability in the presence of low-frequency noise such as earth's magnetic field, and therefore a standard high  $T_c$  SQUID is still being developed.

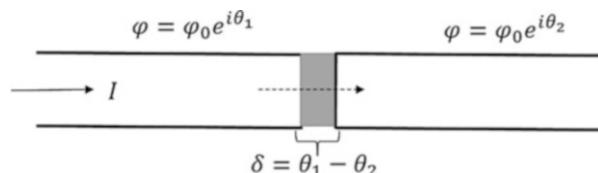
As illustrated in Fig. 6, Josephson's junction is made of two superconducting materials separated by a thin insulating layer. The DC applied current through the junction,  $I$  relate to the phase difference ( $\delta = \theta_1 - \theta_2$ ) of the two superconductors and the critical current  $I_c$  of the junction, according to Eq. (12). The voltage across the junction is obtained by the temporal derivative of the phase difference  $\delta$  when  $I > I_c$ , shown in Eq. (13).

$$I = I_c \sin \delta. \quad (12)$$

$$V = \frac{\hbar}{2e} \left( \frac{d\delta}{dt} \right). \quad (13)$$

A SQUID is formed by a superconducting loop consisting of either one or two Josephson junctions (shown in Fig. 6) that contains flux only in multiples of the flux quantum, i.e.,  $n\emptyset_0$ , where  $\emptyset_0 = 2.07 \times 10^{-15} \text{ Wb}$ . Therefore, a change in the flux

**Fig. 6** Illustration of Josephson junction in SQUID



applied to the loop will cause currents to flow to oppose that change. This causes a phase difference across the junction, which is then recorded as a voltage change in the output. Details of the physics behind SQUID technology and the various types of SQUIDs such as DC SQUIDs and RF SQUIDs are discussed in detail in (Jenks et al. 1997; Weisenstock 1991). Since SQUIDs have been explored extensively in the past and several reviews are available in the literature (Jenks et al. 1997; Tumanski 2013), in this chapter the authors have focused on the most recent advancements in this technology. The first breakthrough of SQUIDs in imaging technology, especially as a biomagnetic NDE system, took place in Biomagnetic Technologies Inc. (BTI) and Quantum Design for John Wikswo at Vanderbilt University (Buchanan et al. 1989). Since then, the capability of high-temperature superconducting (HTS) SQUIDs has grown rapidly starting from aerospace composite structures to medical applications.

## Eddy Current Magneto-Optic Imaging (MOI) Sensors

The eddy current Magneto-Optic Imaging (MOI) uses the Faraday rotation effect, which involves the rotation of the plane of polarization of linearly polarized light while transmission through the magneto-optic material (Lee et al. 2005). Figure 7 illustrates the Faraday rotation effect and the schematic of the MOI instrument with magneto-optic sensor. The light source is typically a LED array, and receiver is a CCD camera. In a homogeneous specimen, the magnetic field due to the eddy currents is tangential to the inspected surface. Anomalies and cracks interact with eddy current producing the magnetic field along the easy axis of the magneto-optic sensor. Distortion in the magnetic domain in the sensor due to the external magnetic field results in the rotation of linearly polarized light by an angle given by Fitzpatrick et al. (1993)

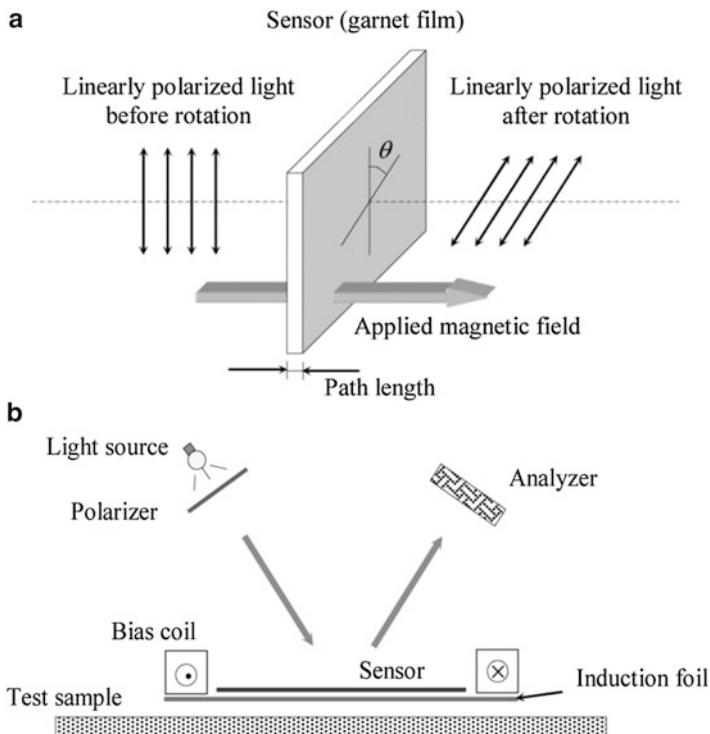
$$\theta = \theta_f(\mathbf{k})l/(|\mathbf{k}||\mathbf{M}|), \quad (14)$$

where  $\mathbf{M}$  is the local state of magnetization,  $\mathbf{k}$  is the wave vector of incident light,  $l$  is the thickness. The Faraday rotation angle  $\theta_f$  can be 3–4° depending on the thickness. The reflected light is viewed through an analyzer in the form of images for dark or light areas corresponding to the strength of magnetization. To adjust for stray DC fields and varying field strengths, a bias coil is used to generate a DC bias field for adjustments (Fitzpatrick et al. 1993; Fitzpatrick et al. 1996). The MOI has been used extensively in aging aircraft inspections, which would be discussed in the next section.

---

## NDE Applications and Beyond

Fluxgate sensors and SQUID sensors can provide sensitivities in the range of pT without sacrificing the spatial resolution. The AMR and GMR sensors provide better spatial resolution, but the field sensitivity and temperature stability are low (Rifai



**Fig. 7** (a) Faraday rotation effect and (b) schematic of magneto-optic imaging (MOI) instrument. (Reproduced from Zeng et al. (2006) with permission from IEEE)

et al. 2016; Nair et al. 2006; Dogaru and Smith 2001). The SQUID sensor has the most desirable properties, but the price due to the cooling requirement is high. Each sensor has its own advantages for which they are used in specific NDE applications. Table 4 outlines a list of low-field NDE applications, the sensor used and defect details.

### Pipeline and Tube Defects Inspection

One of the most popular methods to inspect pipeline rapidly is the magnetic flux leakage (MFL) detection methods (Suresh et al. 2017; Coramik and Ege 2017). A typical MFL tool integrated pipeline inspection gauge (PIG) comprises a drive section at the front of the tool, a central magnetizer section, and a data logger situated at the rear end (Shi et al. 2015). Systems that are more complex have been developed to suit pipelines with curved sections as shown in Fig. 8. Compared to other measurement methods used in the PIG, the Hall sensor is mature with better stability and the temperature characteristics, and thus it is preferred for magnetic field leakage measurement. The coverage and resolution have been increased with

**Table 4** Summary of low-field nondestructive evaluation-based applications and details

Application	Excitation and sensing	Defect type	Defect size	References
Pipeline corrosion damage	Magnet exciter and Hall sensor array	Corrosion	40 mm length and 3 mm deep	Li et al. (2018)
Tube defects	Bobbin induction coil exciter and TMR sensor array	Crack	30 mm length, 2–8 mm deep	Yuan et al. (2017)
Transient ECT of multilayered conducting materials	Double “D” excitation coils with fluxgate sensor	Rectangular subsurface defect	$70 \times 2 \times 1 \text{ mm}^3$	Nagendran et al. (2018)
Thermoelectric fluxgate gradiometer for subsurface defects	Thermoelectric current excitation and fluxgate sensor	Subsurface tin semispherical inclusion in copper	3.18–12.7 mm	Carreon (2006)
Buried objects	Earth’s magnetic field and fluxgate sensor	Buried object	Dia 11.7 cm, Height 4.9 cm	Ege et al. (2016)
Magnetic defects in PMSM	Magnetic field of PMSM and fluxgate sensor	Absence of 7% magnets	—	Goktas et al. (2016)
Wire rope defects	Single-core and single-winding fluxgate for actuation as well as sensing	Broken wire	Number of broken wires: 1–9	Wei and Jianxin (2002)
Crack detection and profiling	Coil exciter and Hall sensors based magnetic camera	Crack	Crack length 5–15 mm, depth 2–9 mm	Jun and Lee (2008)
Composite sandwich structure	Pulsed eddy current coil and Hall sensor	Insert	1 × 1.5 mm	He et al. (2014)
Notch detection in CFRP laminate	Two induction coil exciter and TMR sensor	Notch and Flat bottom hole	10 × 1 × 4 mm and 3–10 mm diameter and 0.8–2.4 mm depth	Rosell et al. (2018)
Impact damage detection in 5 mm thick CFRP samples	Excitation coil and SQUID magnetometer sensor	Impact damage	4–16 J impact	Wang et al. (2017)
Aircraft lap joint	Inductive flat coil exciter and Hall sensor	Void	10 mm	Sophian et al. (2002)
Riveted multilayer joint	Induction coil exciter and MOI sensor	Notch in subsurface layer	2–10 mm	Deng et al. (2006)

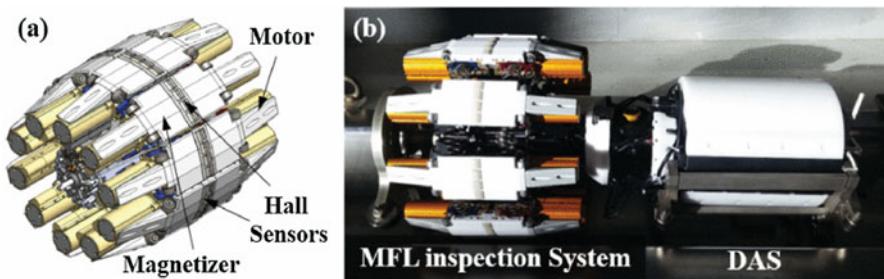
(continued)

**Table 4** (continued)

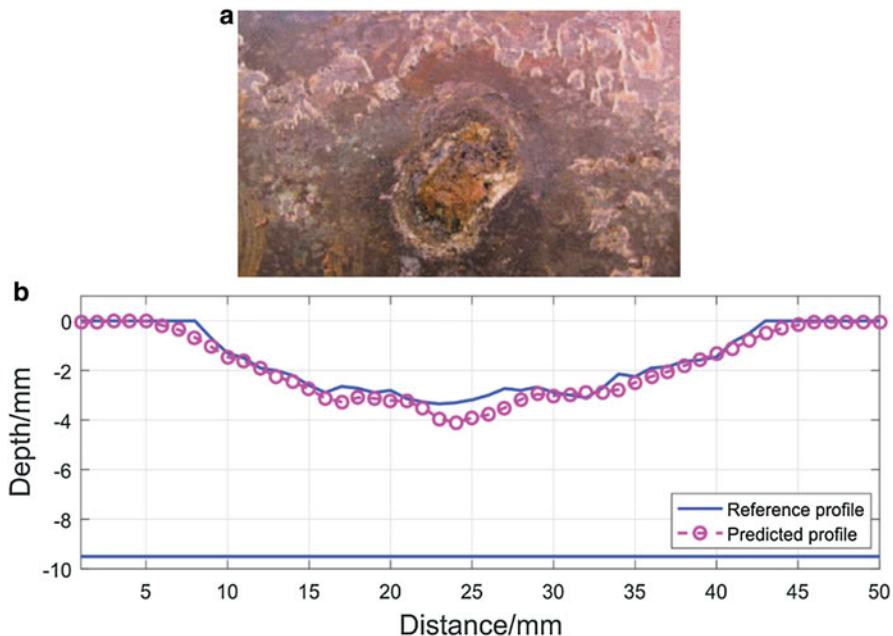
Application	Excitation and sensing	Defect type	Defect size	References
Riveted 3 layer joint	Induction coil exciter and GMR sensor	Notch in middle layer	—	Yang et al. (2015)
Riveted 2 layer joint	Induction coil exciter and TMR sensor	Notch in lower layer	—	Ye et al. (2018b)
Mechanical properties	Spontaneous magnetization due to strain and Hall sensor	Strain	0.7–13.8%	Oota et al. (2001)
Mechanical stress	U-shaped magnetic core, coil and Hall sensor	Stress	140 kN	Zhou et al. (2017)
Plastic strain	Strain-based spontaneous self-excitation and Hall sensor	Plastic strain	2.5–11% plastic strain	Oota et al. (2001)
Rail wheel defects	Hall sensor array	Cracks	Width 0.2–0.8 mm, Depth 2–8 mm, Angle 30°,45°,75°,90°	Chen et al. (2018)
Metrology thickness measurement	Air core coil and Hall sensor	Thickness	0.1–1 mm	Fan et al. (2017)

more sections in the Hall element array and multichannel data acquisition recently. Kim et al. (2018a) reported an accurate scheme to determine the circular defect depth using MFL. The experimental pull-rig test facility had a fabricated MFL system with 120 Hall sensors integrated with the in-pipe robot system. They used a quadratic function of peak amplitude with estimated values about length and width to express the depth.

Li et al. (2018) designed a signal acquisition device and used Hall sensors to detect defects like natural corrosion in a pipeline. A total of 72 Hall sensors were arranged in 12 capsules around the circumference giving a resolution of 7.5 mm. Detection speed was limited to 3 m/s for satisfactory data acquisition. The pipeline outer diameter and the normal wall thickness were 219 mm and 9.5 mm, respectively. The actual image of the natural corrosion considered is shown in Fig. 9(a). The magnet used in the experiment was N48 with a coercive force of 876,000 A/m and residual magnetism of 1.37–1.43 T. The converted magnetic-to-electrical signals from Hall sensor while scanning the corrosion specimen were recorded using the MFL inspection tool. They proposed a modified harmony search algorithm for quick and accurate reconstruction of defect profiles. Figure 9(b) shows that the predicted profiles fit the reference profiles perfectly. Lu et al. (2017) have reported an advanced



**Fig. 8** A magnetic flux leakage inspection system employing Hall sensors for inspection of pipes with curved sections. (a) 3-D design showing the sensor, motor, and magnetizer. (b) Manufactured system based on the 3-D design. (Reproduced from Kim et al. (2018b) with permission from IEEE)



**Fig. 9** (a) Corrosion defect in a pipeline. (b) Reconstruction result of the natural corrosion defect using data from Hall sensors post-processed by a modified harmony search algorithm showing the reference and predicted profiles in terms of the measured depth. (Reproduced from Li et al. (2018) with permission from IEEE)

reconstruction of arbitrary defect profiles considering the velocity effect due to the PIG motion. Recently, Piao et al. (2019) reported a novel technique for fast reconstruction of a 3-D defect profile from MFL signals by combining key physics-based parameters and data-driven machine-learning algorithms. Local and global feature

parameters of the defect are determined using nonlinear least squares approach and a rational Bezier curve model generates the 3-D defect profile.

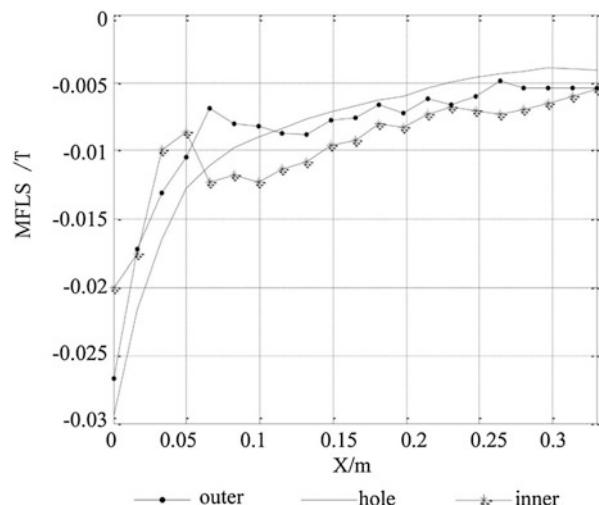
Liu et al. (2015a) developed an advanced algorithm based on analysis of Hall sensor data using magnetic dipole theory to distinguish between inside and outside defects in pipelines. Hall sensors were used to detect the leakage fields due to metal loss. The inner-outer defect's length was 30 mm with width 10 mm and depth of 25% of the wall thickness. A clear distinction between the inner and outer defects in the radial signals was observed as shown in Fig. 10, whereas the axial signals were identical. This is applicable to a weak magnetic field where the material inherent magnetic moment is considerable with a measurable difference in the distribution of radial components of the inner and outer defect. Under a strong magnetic field, the resistance effect of the atomic magnetic moment of ferromagnetic material is negligible.

Wu et al. (2017) reported a further improvement to the detection methods using Hall sensors by superimposing alternating magnetic field on the MFL magnetization field. The angle between the crack and the magnetization direction is called azimuth angle  $\beta$ . As the crack or the magnetization field  $B$  is tilted, the flux direction near the crack changes. Magnetic flux density  $B'$  across the crack walls is given by  $B' = B\cos(\beta)$ . The principle of magnetic dipole moment gives the magnetic charge density on the crack walls. In the presence of an additional alternating magnetic field, the excitation generated in the new structure is given by

$$\phi = \phi_{MFL} + \Delta\phi = \phi_{MFL} + K \cdot \phi_{MFL} \sin(2\pi ft), \quad (15)$$

where  $\phi_{MFL}$  is an excitation from permanent magnets,  $\Delta\phi$  is the alternating field excitation, and  $K$  is a coefficient varying between 0.2 and 0.4. The second part due to the alternating field generates cross-sectional eddy currents in a circumferential direction. This eddy current field forms perpendicular to the magnetization direction

**Fig. 10** Radial component of magnetic flux density measured using a Hall sensor in a pipeline with metal loss type of defect. (Reproduced from Liu et al. (2015a) with permission from Elsevier)



in the pipeline wall. The defects lead to an additional disturbance of the circumferential eddy current field. A signal processing algorithm was proposed to estimate a complex variable  $T$  representing the circumferential and axial signals as real and imaginary components, respectively, as

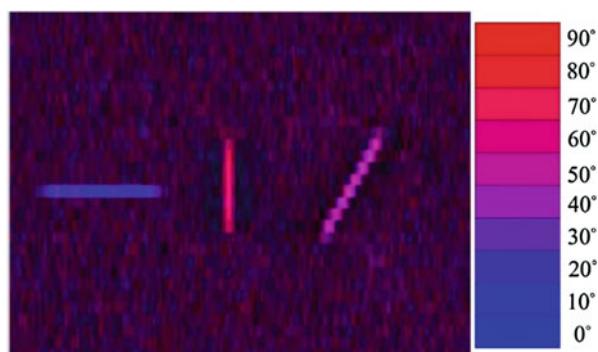
$$T = B_{MFL} + jE|\tilde{B}_{IMF}|, \quad (16)$$

where  $B_{MFL}$  represents the circumferential detection results,  $E$  is the balance factor, and  $\tilde{B}$  represents the axial detection results. Figure 11 shows the detection of differently oriented cracks possible through RGB composition using real (Red) and imaginary (Blue) parts of the complex signal  $T$ . The detection results for circumferential cracks are clearly visible in the real part, whereas the longitudinal (axial) cracks are vague. The imaginary part illuminates the longitudinal cracks quite clearly with the transverse cracks hardly seen. Thus, the synchronous inspection in orthogonal direction was possible without missing the crack with different orientations.

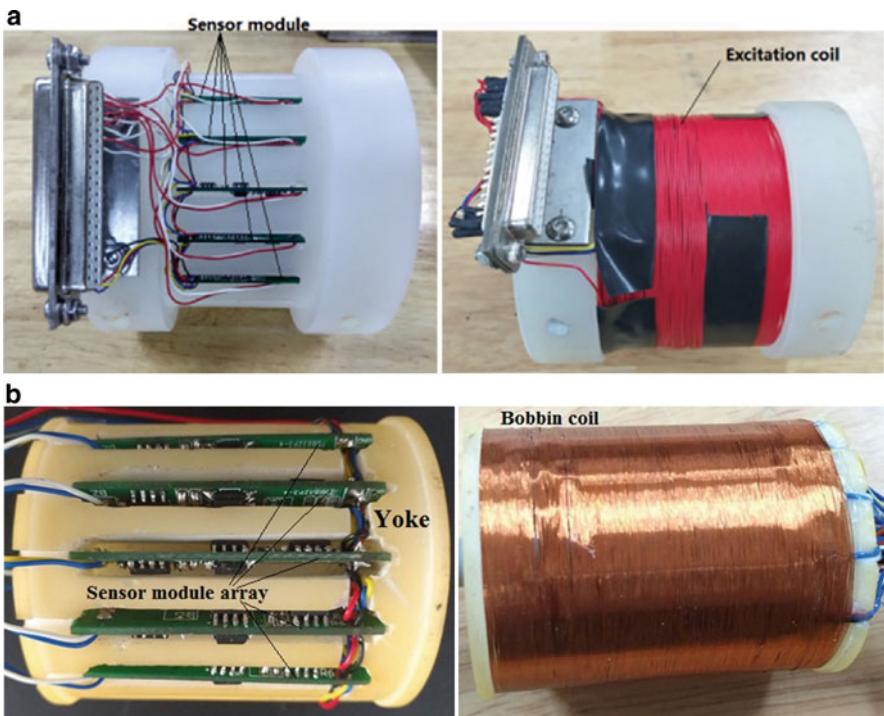
The scan velocity has a significant influence on the detection capability. Advanced reconstruction of arbitrary defect profiles considering the velocity effect due to PIG motion has been reported recently. Pullen et al. (2018) found a signal reduction of up to 53% due to the increase in scanning velocity from 0.5 to 3 m/s. Utilizing the small dimensions and large range, Suresh et al. (2017) used Hall sensors to develop an MFL system for detecting a defect in small diameter steam generator tubes. They proposed magnetizer arrangement with the capability to produce a uniform magnetic field with which they were able to detect a 40% depth defect in 5 mm flat bottom hole (FBH).

Power plant tubes and pipes involve inspection of inner or exterior walls of steam generator tubes, steam piping, reheat steam piping, feed and cooling piping, and auxiliary piping. These parts make most of the surface area required for heat and mass transfer. Processes like electrochemical corrosion, thermal fatigue at high temperatures, and high-pressure stress cause initial longitudinal fatigue cracks that begin and grow rapidly in these areas. Ultrasonic and MFL techniques are not sensitive to narrow fatigue cracks, and ECT is sensitive to lift-off and coatings

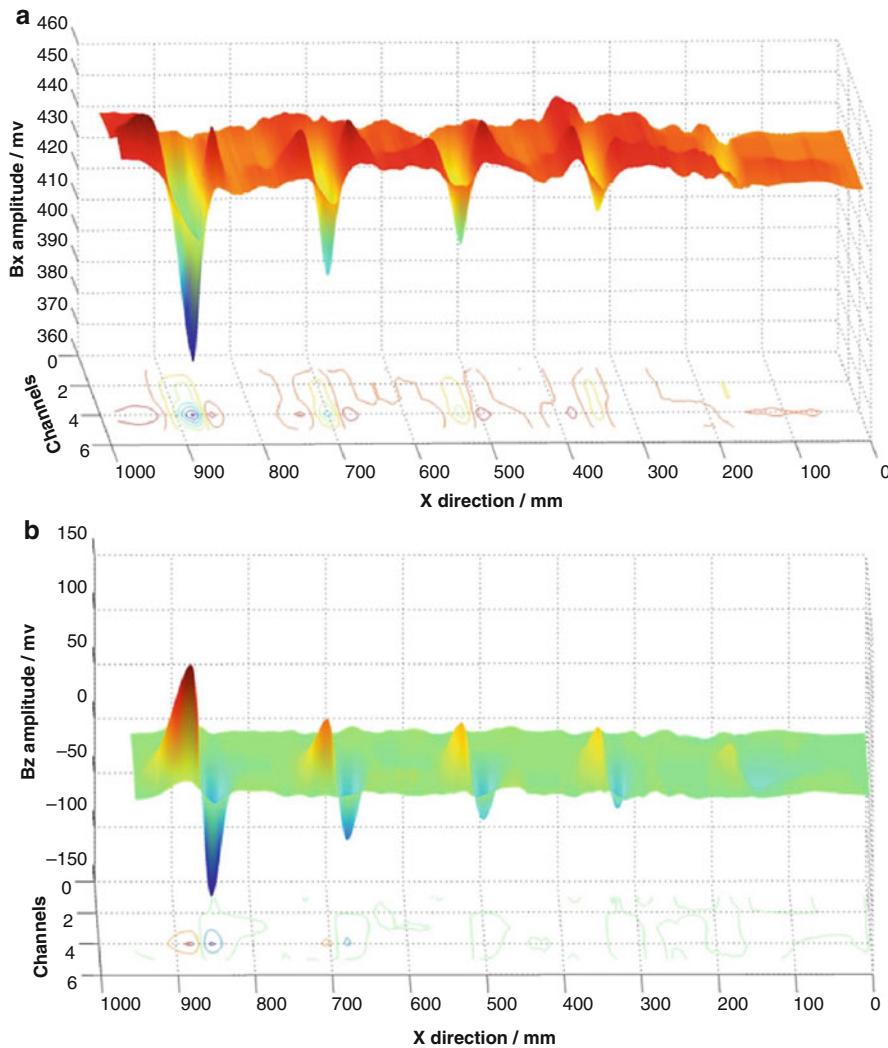
**Fig. 11** Color image of the scanning results in real and imaginary parts of the complex variable output estimated using the signals obtained from fluxgate sensors by scanning a pipe specimen with cracks of different orientations  
(Reproduced from Wu et al. (2017) with permission from Elsevier)



(Afzal and Udpal 2002; Zhou et al. 2016). Muñoz et al. (2016) used the magnetic perturbation method with GMR-based probes to evaluate the length, width, and orientation of cracks from the peak value of signals and distance between the peaks in the scan plot. Following a similar approach, Yuan et al. (Yuan et al. 2017, 2018) used current field perturbation method (Li et al. 2016) to detect outer and inner surface cracks and defects. A coaxial encircling excitation coil is used to induce the circumferential current on the sinusoidal excitation. The entire inner and outer circumference is scanned by two types of probes consisting of distributed TMR sensor arrays as shown in Fig. 12. Orthogonal TMR sensors measure the magnetic field in the axial  $B_x$  and the normal direction  $B_z$ . Figure 13 shows the response of sensors when the probe passes around the stainless steel tube (outer dia. 65 mm and inner dia. 47 mm) with 30 mm length crack of different depths, 2 mm, 4 mm, 6 mm, 8 mm, and a through crack. A linear variation of signals  $B_x$  with a linear relationship versus depths was shown through a suitable calibration-enabled measurement. The location of opposite peaks in the  $B_z$  scan with suitable calibration indicated the length of the crack.

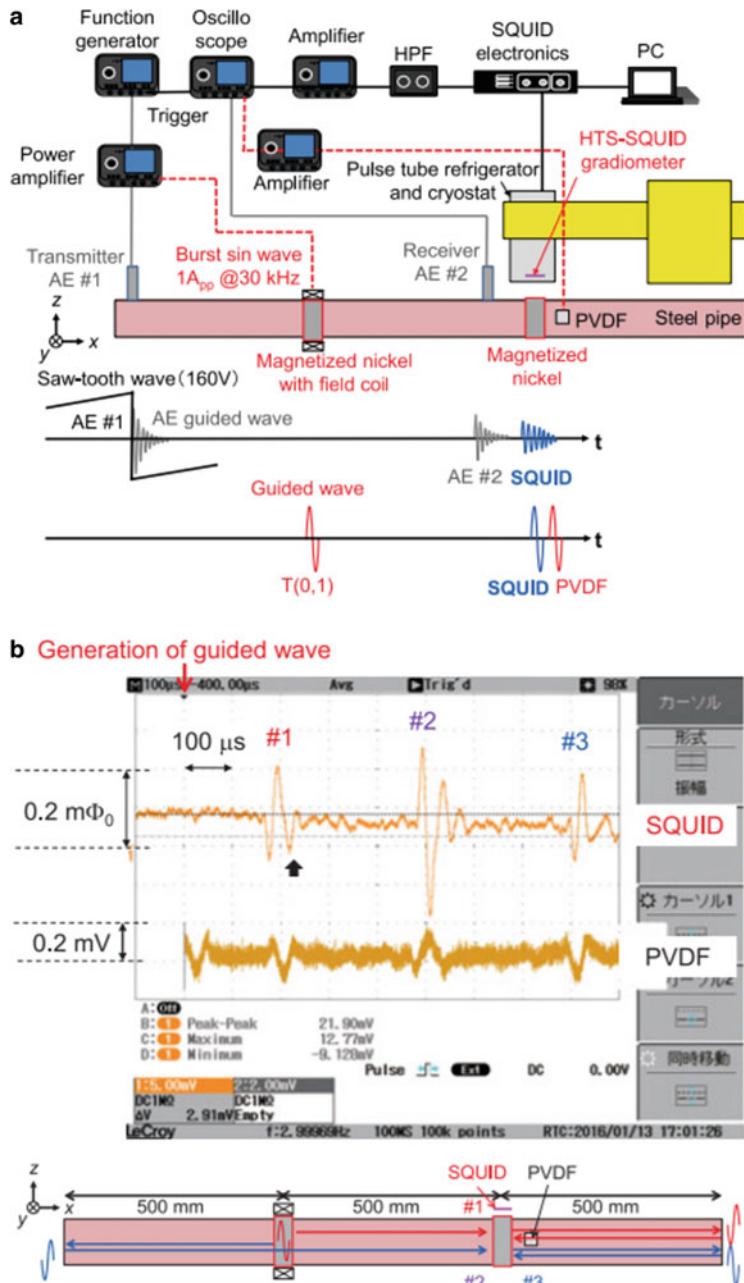


**Fig. 12** Probes for circumferential current testing system for tube inspection with TMR sensor module array in a yoke (left) and a bobbin coil for excitation (right). **(a)** For external inspection. (Reproduced from Yuan et al. (2017) with permission from Elsevier). **(b)** For internal inspection. (Reproduced from Yuan et al. (2018) with permission from Elsevier)



**Fig. 13** Magnetic field response (a)  $B_x$  and (b)  $B_z$  from TMR sensors due to cracks on stainless steel pipe with a crack of length 30 mm and depth 2 mm, 4 mm, 6 mm, 8 mm and a through wall crack. (Reproduced from Yuan et al. (2017) with permission from Elsevier)

Behun et al. (2018) reported the ability to estimate the spatial orientation, width, and length of the cracks by estimating the wavelet transform of the GMR signals and using artificial neural networks. They considered artificial notches of different dimensions and established an efficient algorithm for fast computation. Arismendi et al. (2018) presented the classification of notches based on width and orientation using GMR and fluxgate sensor data. Bailey et al. (2017) studied the evaluation of corrosion type of defects undercoating with GMR sensor and a pipe encircling



**Fig. 14** (a) Schematic diagram of the SQUID NDE system for pipes and images of transmitted AE and ultrasonic guided waves. (b) Measurement of ultrasonic guided wave in T(0, 1) mode by HTS-SQUID gradiometer (upper waveform). (Reproduced from Hatsukade et al. (2017) with permission from IEEE)

exciter (300A, 2–500 Hz). A linear correlation with defect volume and response from  $1.2 \times 10^{-3}$  to  $1.6 \times 10^{-3} \mu\text{T}/\text{mm}^3$  was seen. Sharatchandra Singh et al. (2015) reported a 12-element flexible GMR array probe to scan thermal power plant tubes for defects such as flat bottom holes, erosion, and notches. The flexibility of the probe allowed it to conform to any diameter tube. Sasi et al. (2018) reported the use of two orthogonally oriented GMR sensors and an excitation coil to detect 2–7 mm deep notches.

Hatsukade et al. (2017) have investigated the use of high-temperature SQUIDs for investigation of deep-rooted defects in aluminum alloys (Hatsukade and Tanaka 2016) as well as defect detection in nickel pipes combining SQUID and ultrasonic guided wave sensors. Figure 14(a) depicts a schematic of the SQUID NDE system implemented for pipeline investigation where the guided wave was generated by the magnetostrictive transmitter and received by a SQUID gradiometer. As seen in Fig. 14(b), the SQUID receiver shows a much higher SNR compared to a polyvinylidene difluoride (PVDF) piezoelectric film, which is usually used in a typical guided wave system.

## Subsurface Inspection

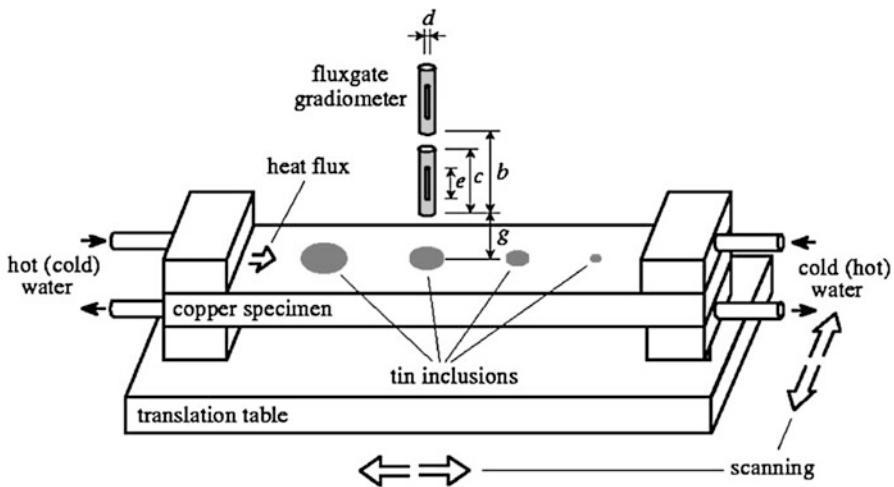
Inclusions in materials are difficult to detect if they have similar mechanical and electric properties. Thermoelectric power is sensitive to such small change in material imperfections. The volumetric effect related to electron diffusion kinetics and non-homogeneous contact enhances this effect (Carreon et al. 2000). Due to an externally induced temperature gradient, the imperfections form thermocouples and produce thermoelectric currents. The surrounding material serves as the reference electrode and sensitivity to subtle effects causing imperfection can be high. The magnetic field produced by thermoelectric current around an inclusion is given by Carreon et al. (2000):

$$H_0 = -a\nabla T\sigma(S' - S)\Gamma, \quad (17)$$

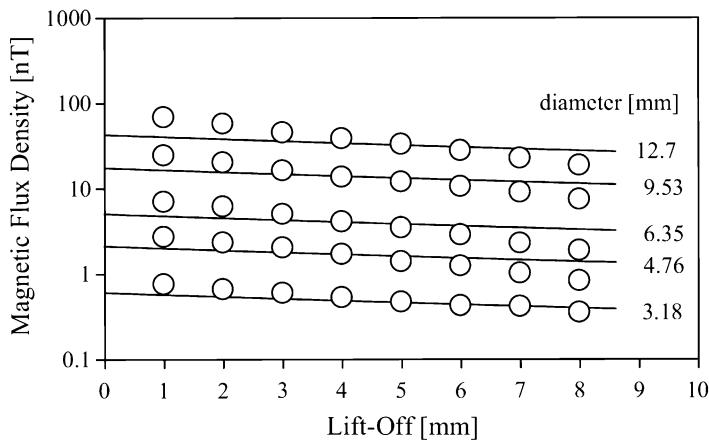
where  $a$  is the size of the inclusion,  $\nabla T$  is the temperature gradient,  $\sigma$  is the electric conductivity of the host, and  $S' - S$  is the relative thermoelectric power of the inclusion. The scalar constant of the magnetic field strength  $\Gamma$  for a spherical inclusion is given by Carreon (2006) as:

$$\Gamma = \frac{3}{\left(1 + 2\frac{\sigma}{\sigma'}\right)\left(2 + \frac{K'}{K}\right)}, \quad (18)$$

where  $K$  is the thermal conductivity of the host. Carreon (2006) used Hall sensors to detect the magnetic flux density due to these currents arising from tin inclusions in a copper bar. The schematic of the experimental setup used is shown in Fig. 15. Spherical holes with 3.18–12.7 mm in diameter were milled in a copper plate with a size of 12.7 mm × 38.1 mm × 500 mm. Molten tin was poured to form spherical inclusions. A modest temperature gradient of 0.7 °C/cm was obtained from the flow of cold and hot water through the specimens that produced a flux density of



**Fig. 15** Schematic of experimental arrangement to create thermoelectric effect in a copper bar with tin inclusion using temperature gradient created by hot and cold flowing water. (Reproduced from Carreon et al. (2000) with permission from AIP)



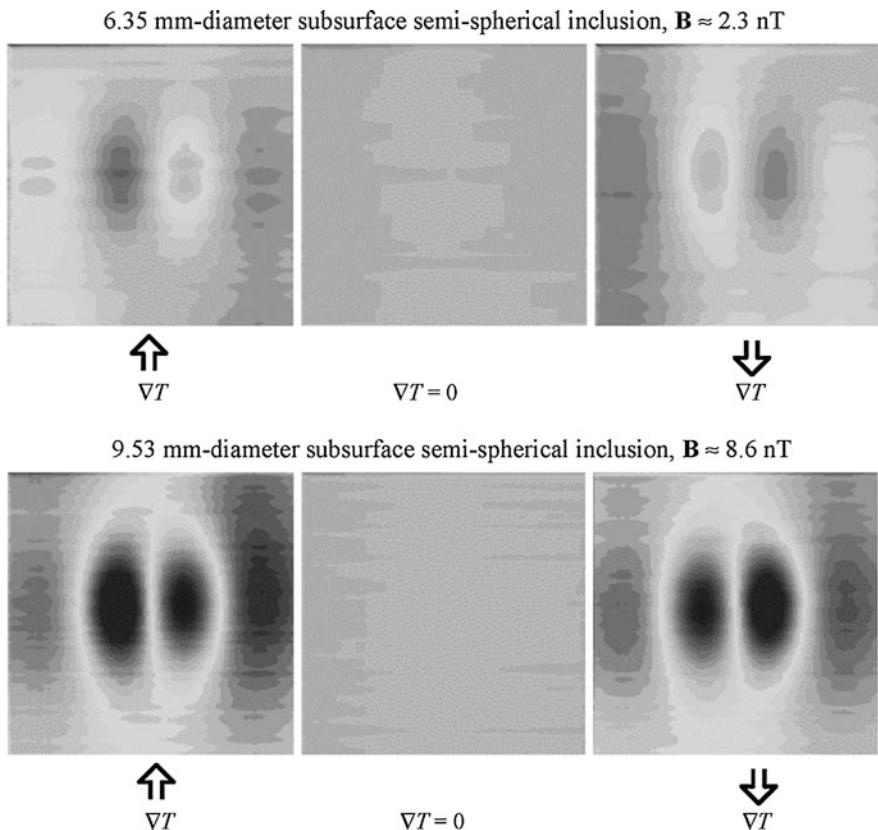
**Fig. 16** Experimentally measured (solid circular marker) and theoretically predicted (solid line) peak-to-peak magnetic flux densities variation with lift-off distance form subsurface semispherical tin inclusions. (Reproduced from Carreon et al. (2006) with permission from AIP)

1–250 nT. The specimen was scanned with a pair of magnetic fluxgate sensors having a sensitivity of 10 mT/V, forming a gradiometer arrangement.

Figure 16 shows the change in peak-to-peak magnetic flux density with the lift-off distance varying from 1 mm to 8 mm and for five inclusions of diameters ranging from 12.7 to 3.18 mm. The solid lines corresponding to the analytical results from

Eq. (18) match the solid points representing the experimental results. The small changes in the slope sign were different from that expected due to the assumption that the specimen is infinitely wide. Figure 17 shows the magnetic field images recorded at 2 mm distance above the copper specimen surface with a fluxgate sensor showing the two subsurface semispherical tin inclusions. Due to the change in direction of the temperature gradient in the specimen, the bipolar lobes change sign. With the increase in the lift-off distance, the lobes get wider and the magnitude of the magnetic flux density decreases. In the case of subsurface inclusions, the magnetic field became significantly weaker and distortions in the spatial distribution of the field can be seen.

Conducting plates can have subsurface defects at various depths. Nagendran et al. (2018) reported the detection of defects in aluminum plates by using a fluxgate sensor in a combination of double “D” differential coil to enhance the depth of investigation. The schematic of the setup used in their experiments is shown in



**Fig. 17** Peak magnetic flux density distribution images obtained from a fluxgate sensor by scanning a copper specimen with subsurface semispherical tin inclusions. Thermal gradient  $\nabla T \approx 0.7 \text{ }^{\circ}\text{C}/\text{cm}$ , 2 mm lift-off distance, 76.2 mm  $\times$  76.2 mm scanning dimension. (Reproduced from Carreon (2006) with permission from Elsevier)

**Fig. 18** Schematic of an NDE system involving transient eddy currents with fluxgate sensor for the detection of the crack type of defects. (Reproduced from Nagendran et al. (2018) with permission from Springer)

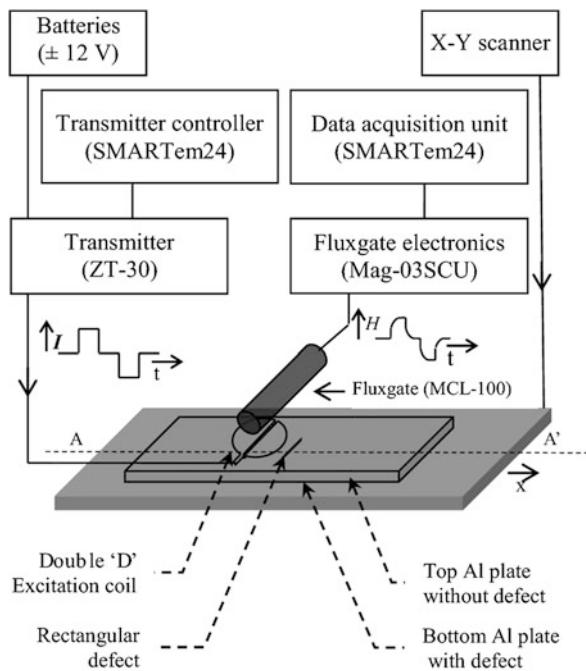
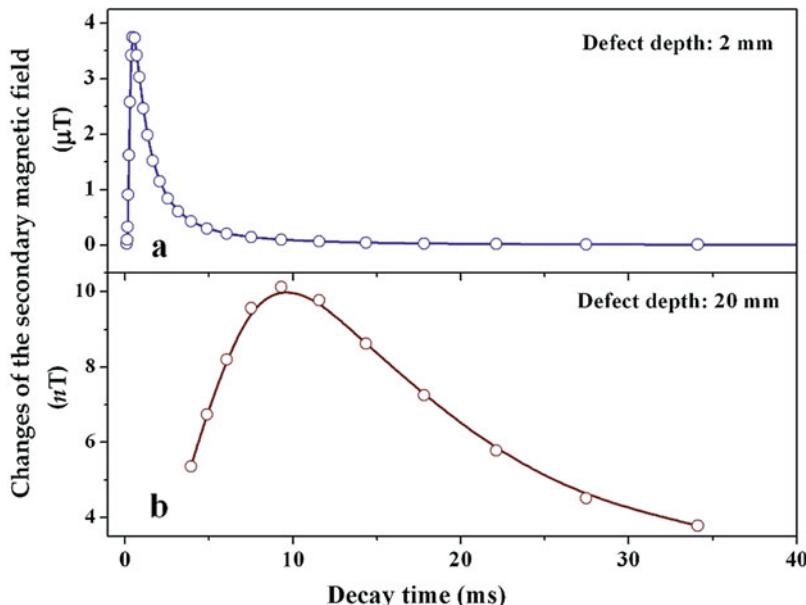


Fig. 18. The fluxgate measured the magnetic field directly and the double “D” coil reduced the direct magnetic field coupled to the fluxgate sensor, thereby giving an opportunity to apply a large magnetic field to increase the induced eddy current. Aluminum plates  $300 \times 200 \text{ mm}^2$  with a thickness ranging from 2 to 20 mm were considered. A plate of thickness 1 mm and defect size of 70 mm length, 1.5 mm height, and 1 mm width was considered. The defect plate was located at 2 mm and 20 mm below the stack of aluminum plates. A trapezoidal pulsed current of 1 A with a repetition rate of 400 ms was used for excitation. Figure 19 shows the changes in the measurement of the fluxgate sensor output due to a defect at a depth of 2 mm and 20 mm for different decay times. The defect depth of 2 mm results in a maximum change at an earlier time gate of 0.56 ms. For the same defect at a depth of 20 mm, the maximum change results at a later time gate of 9.4 ms. The decay time gives information about the depth of the subsurface defect directly.

## Defects in Permanent Magnets

Permanent magnet synchronous motors (PMSMs) have high efficiency, torque/current ratio, and power density, reasons for which industries likely prefer them. Damages like broken magnets and demagnetization are commonly limiting their use due to maintenance and safety concerns (Espinosa et al. 2007; Prieto et al. 2011; Urresty et al. 2013). Drawbacks like size, noise, and installation problems arise when



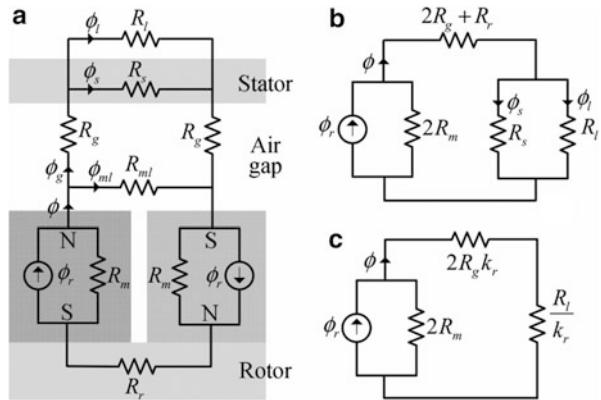
**Fig. 19** Changes observed in the secondary magnetic field measured by a fluxgate sensor at different decay times due to defect depth (a) 2 mm below and (b) 20 mm below the surface. (Reproduced from Nagendran et al. (2018) with permission from Springer)

using induction coils to measure the leakage flux due to permanent magnet defects. Goktas et al. (2016) proposed a remote and on-case monitoring solution to measure the magnetic flux leakage using fluxgate sensors and a conditioning circuit to detect defects in the permanent magnets. They analyzed the characteristic harmonics by simulations using a stepping finite element method. The leakage flux on the stator winding and the motor structure is due to the leakage current, back-EMF, torque, and stator flux (Urresty et al. 2013). The PMSMs generate strong magnetic fields up to 1 Tesla in the air gap. A magnetic equivalent circuit as shown in Fig. 20 was introduced with all possible reluctances like rotor  $R_r$ , stator  $R_S$ , air  $R_g$ , and magnet  $R_m$  to establish the relationship of the leakage flux given by

$$\mathcal{O}_l = \frac{k_l \mathcal{O}_r}{1 + k_r \frac{\mu_r}{P_{cg}} + \frac{\mu_r}{4k_r P_{cl}}}, \quad (19)$$

where  $k_l$  and  $k_r$  are the leakage and reluctance factors, respectively,  $P_{cg}$  and  $P_{cl}$  are the permanent flux coefficients, and  $k_{mi}$  is the magnetic leakage factor. The possible failures create characteristic harmonics in the air gap flux that are reflected in the leakage flux. Disturbance due to off-centered rotor and supply voltage can affect the leakage flux as well. An integrated fluxgate sensor (DRV 421) was used that has a very small footprint of  $4 \times 4 \text{ mm}^2$  with integrated compensation coil driver and signal conditioning circuitry.

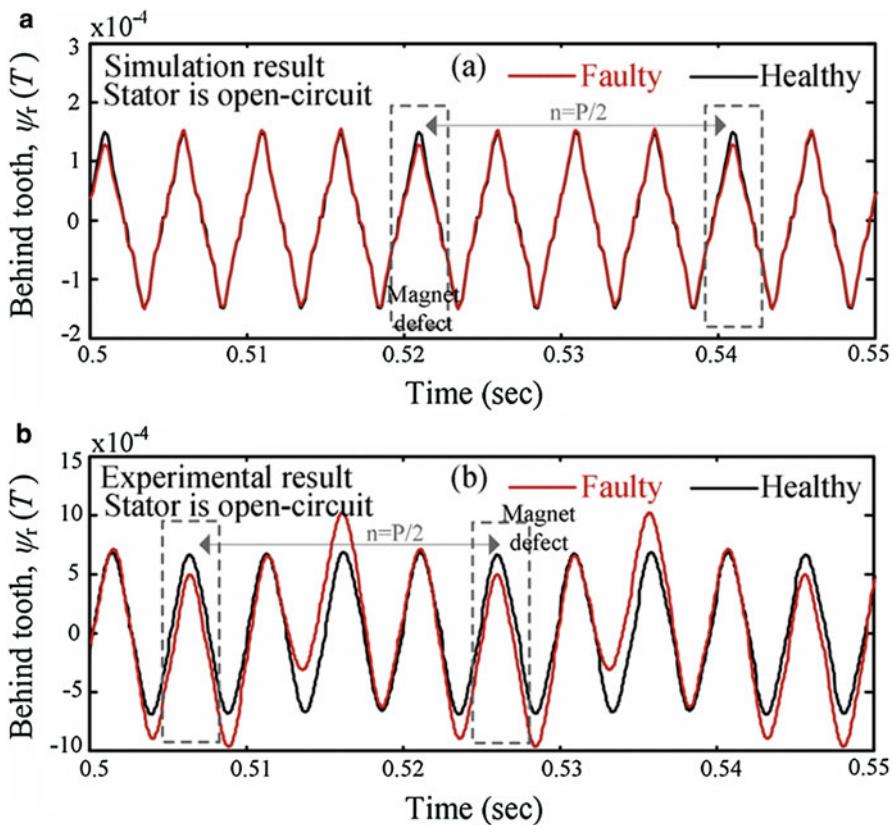
**Fig. 20** Magnetic equivalent circuit of PMSM showing the reluctances and magnetic flux components (a) leakage flux crossover (b) eliminating magnetic leakage flux (c) eliminating reluctances. (Reproduced from Goktas et al. (2016) with permission from the IEEE)



Simulation and experimental results were analyzed in the time domain with the stator in open circuit. This decouples the effects of current and magnets on the leakage flux. The radial ( $\psi_r$ ) and tangential ( $\psi_t$ ) leakage flux components measured on the motor are shown in Figs. 21 and 22, respectively. The radial flux measured behind the tooth and the tangential flux measured behind the slot were obtained at a 200 Hz supply frequency. Due to the magnet defect, both simulation and experimental results show a reduction of leakage flux across the defective magnet. The decreased flux is observed at every mechanical cycle of  $p/2$ , where  $p$  is the number of poles. The leakage flux also becomes asymmetric and was seen particularly at adjacent fluxes of the defected magnets. Several tests confirmed the ability to monitor the magnet defect at one or multiple locations with minimum signal processing support with only one or more rotor cycles. The fluxgate sensor provided a very small footprint with a resolution of the order of nT. In addition, it provided absolute flux values in both frequency and time domain, the ability for remote sensing, and independence of speed and torque on fault detection.

## Buried Features

Buried pipelines and structures are difficult to maintain due to limited access. They require frequent maintenance due to corrosion and other damage caused by the surrounding humidity, temperature, and soil movement. Remotely detection methods are popular in the military, civil, and health areas. Conventional ground-penetrating radar (GPR) (Orlando and Slob 2009), electromagnetic induction spectroscopy (Riggs et al. 2001), infrared visualization (Han and Park 2004), thermal detection (Meola 2007), nuclear quadrupole resonance (Ostafin and Nogaj 2007), nuclear methods (Miri-Hakimabad et al. 2008), and X-ray methods (Birocsa et al. 2009) are high cost techniques and have high power consumption. Inspired by magnetic anomaly detection, Ege et al. (2016) have demonstrated detection of buried materials using a high sensitive fluxgate sensor and the earth's magnetic field to minimize power consumption. They detected samples in soil (with 150  $\mu\text{m}$  average soil particle size) containing nonmagnetic and ferromagnetic

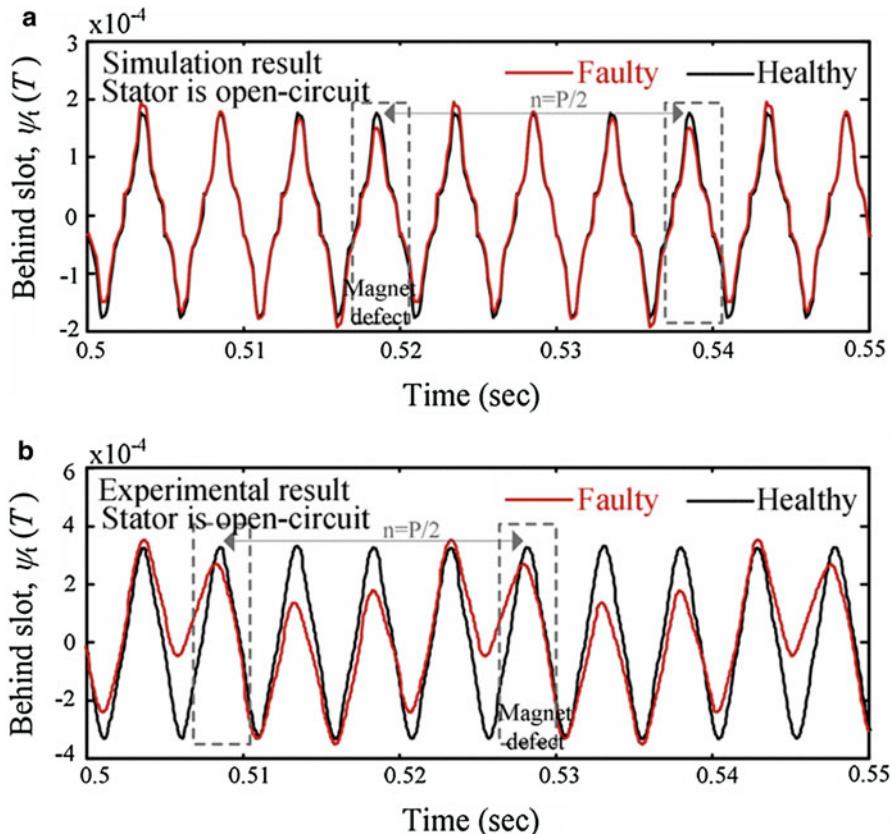


**Fig. 21** Radial flux from a healthy and faulty motor in time domain obtained from the measurements using a fluxgate sensor located behind the tooth, while the stator is in open circuit (a) simulation result and (b) experimental result. (Reproduced from Goktas et al. (2016) with permission from the IEEE)

minerals as shown in Fig. 23. The depth of detection was 20–35 cm deep. They could detect the location, burial depth, burial angle, and sample dimension. This capability holds considerable promise to detect defects in buried structures especially using powerful excitation sources and sensing systems like AMR and GMR.

### Damages in Wire Ropes

Wire ropes used in cranes, lifts, bridges, cable cars etc. are heavy-duty and need a periodic inspection to ensure safety. The leakage field method is the most effective one in detecting common damage and defects like broken wire, corrosion, and wear (Weischedel and Ramsey 1989). Apart from induction coils, Hall sensors are more often used to detect the leakage field. The variation of range of the induction coil

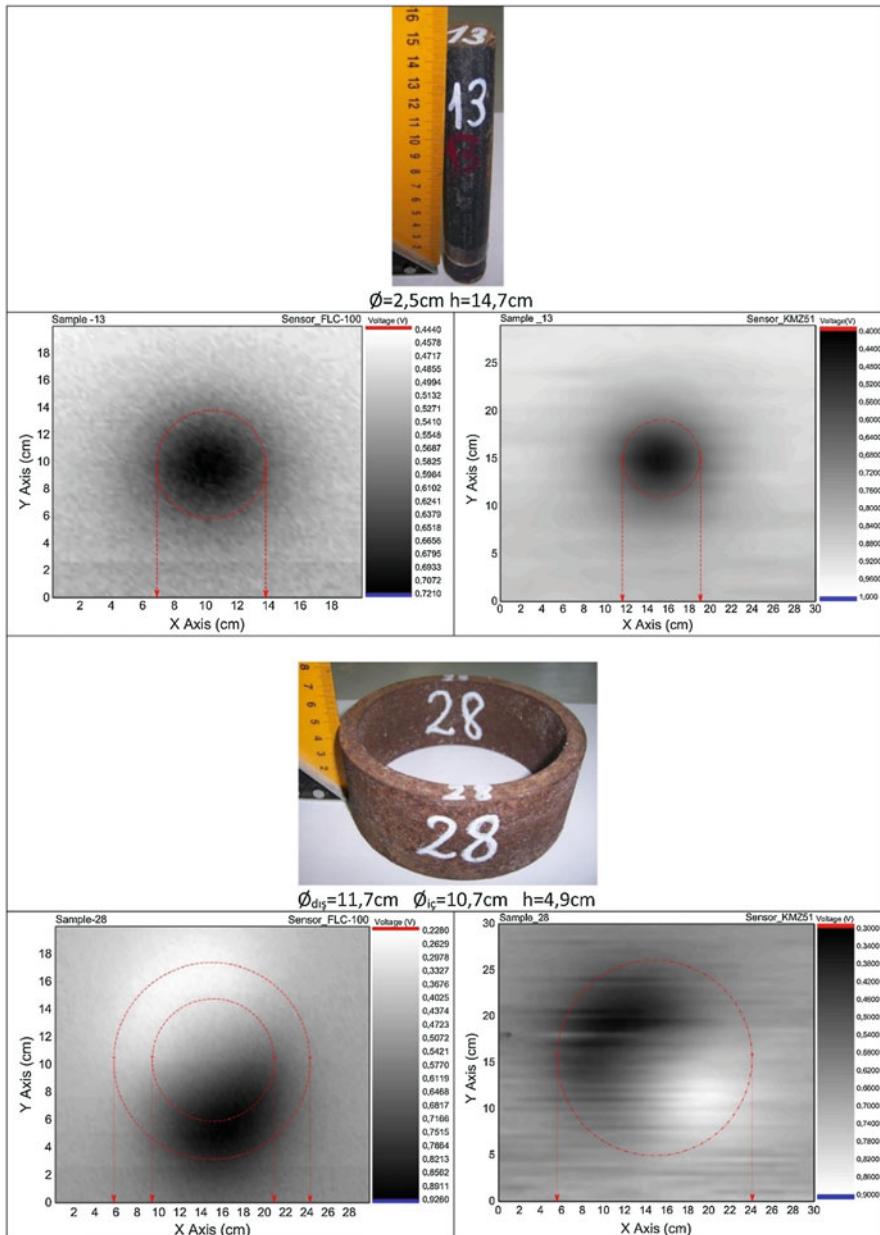


**Fig. 22** Tangential flux from a healthy and faulty motor in time domain obtained from the measurements using a fluxgate sensor located behind the tooth, while the stator is in open circuit (a) simulation result and (b) experimental result. (Reproduced from (Goktas et al. 2016) with permission from the IEEE)

with testing speed and its size is not a proper choice for infield testing. Hall sensors acquire considerable noise and the measurements are affected by temperature. The complexity arises in the distribution analysis due to the wire construction and usage. The leakage field  $\vec{B}_0 = \vec{B}_t + \vec{B}_r + \vec{B}_s$  is a variable spatial vector conveniently resolved in axial  $\vec{B}_t$ , radial  $\vec{B}_r$ , and tangential  $\vec{B}_s$  directions. With  $\vec{B}_0$  being negligible, defects like wire breakage, crack, and corrosion will have the resolved leakage flux along the rope ( $x$ -axis) given by

$$B_t = B_{ta} \sin \left( A_t x + \frac{\pi}{2} \right) + B_{to}, \quad + h < x < h, \quad (20)$$

$$B_r = B_{ra} \sin (A_r x + \pi), \quad + h < x < h, \quad (21)$$

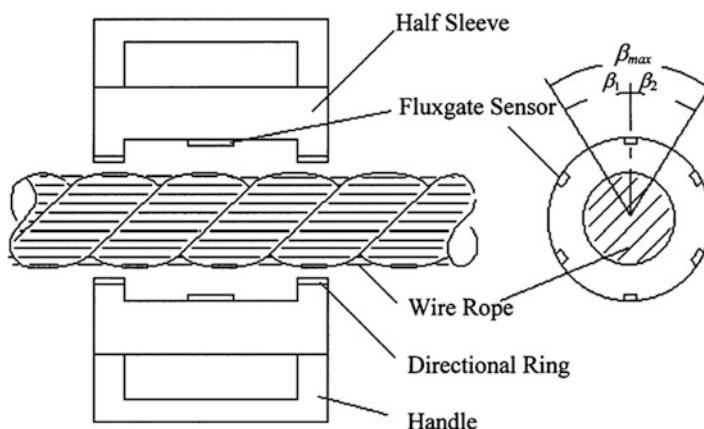


**Fig. 23** Ground surface parallel geometry of the buried samples detected through the soil surface with fluxgate sensor and earth's magnetic field as a source. (Reproduced from Ege et al. (2016) with permission from Elsevier)

where  $B_{ta}$  and  $B_{ra}$  are the amplitudes and  $A_t$  and  $A_r$  are the spatial variabilities.  $B_{to}$  is the axial flux density of magnetization without defects. Both  $B_{ta}$  and  $B_{ra}$  have the essential information about the defect. The flux density of the leakage field varies from 1 to 10 mT and 1 to 5 mT in the axial and radial direction, respectively. The magnetic leakage field inside a wire rope tester lies in the range of 1–10 mT. Therefore, to detect the defects with good accuracy, the sensors should have the range from about 0.1 to 10 mT. Kaur et al. (2018b) have discussed the available sensors and selection criteria for the usage in wire rope testers especially in the context of Hall sensors. Tse et al. (2011) have proposed a solution to sense the magnetic field in wire ropes of different diameter using a flexible magnetostrictive sensor made of a flexible printed film. No loss in performance was detected in the experiments performed.

### Single-Core Single-Winding Fluxgate Sensors

Conventional fluxgate with three windings has a range of 0.01–0.1  $\mu\text{T}$ . However, due to its construction, it is not suitable to test leakage fields from a defective wire. Wei and Jianxin (2002) presented a single-core, single-winding fluxgate sensor, which is a type of variable inductance type of parametric amplifier, making it more sensitive with high SNR than the inductive coil or Hall sensors. Figure 24 shows a schematic of the transducer design. The single-winding in the fluxgate is used for actuation as well as sensing. Thus, the size is reduced to 2–3 mm in width, which meets the requirement to detect wire defects in a wire rope. A square wave excitation voltage with a width or each pulse duration  $\theta$  is applied as the excitation source. The flux density is nonlinear according to the magnetization curve. The relevant information of the defect was obtained by testing the secondary harmonic current  $I_2$  in the winding. The quantitative relationship obtained is given by



**Fig. 24** Schematic of a single-core, single-winding fluxgate transducer used for testing wire ropes for defects. (Reproduced from Wei and Jianxin (2002) with permission from the IEEE)

**Table 5** Detection data of the broken wire type of defects obtained from fluxgate sensors with single-core and single-winding. Reproduced form Wei and Jianxin (2002) with permission form the IEEE

Number	Number of broken wire	Non-error detection times	Correct detection times with $\pm 1$ silk error	Correct detection times with qualitative accuracy
1	1	7	8	7
2	2	8	8	9
3	4	8	9	10
4	7	7	9	10
5	9	6	7	10
Average accuracy		72%	82%	92%

Note: Qualitative accuracy shows whether there is a defect or not

**Fig. 25** Broken strand type of defect in a wire rope made of steel strands. (Reproduced from Kaur et al. (2018a) with permission from Springer)

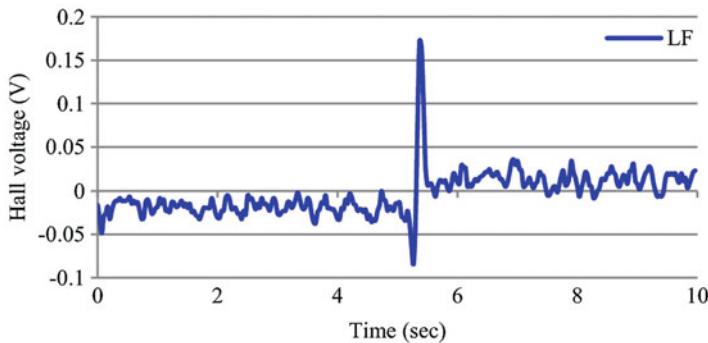


$$I_2 \approx \frac{2l}{\pi N} \sin\theta \left( \frac{1}{\mu'} - \frac{1}{\mu_d} \right) B_0, \quad (22)$$

where  $\mu_d$  is the permeability of the unsaturated core,  $\mu'$  is the saturation segment slope of the magnetization curve,  $B_0$  is the leakage field,  $l$  is the length of the core, and  $N$  is the number of turns in the winding. Using the designed transducer, tests of a wire rope with a defect were carried out by moving the transducer along the wire rope, and the testing data were processed to display the defective state. Table 5 shows the detection data of the wire ropes.

### Parallel Magnetized Magnet Segments and Hall Sensors

Hall effect sensors are also a good choice for wire rope testing. The advantages are good accuracy, robustness, good detection range, easy calibration, and relatively low cost. Radially magnetized ring magnets are widely used, which have difficulty in achieving ideal magnetization. They also require high magnetizing currents (30 kA) and pose logistic challenges (Riley et al. 2000). Kaur et al. (2018a) reported parallel magnetized NdFeB magnet segments used along with Hall sensors to detect local faults and damage like loss in the metallic area and broken strands. The Hall sensor had a sensitivity of 90 mV/mT and a range of 0–10 mT. To detect broken wires as shown in Fig. 25, the Hall sensor is placed perpendicular to the radial component of



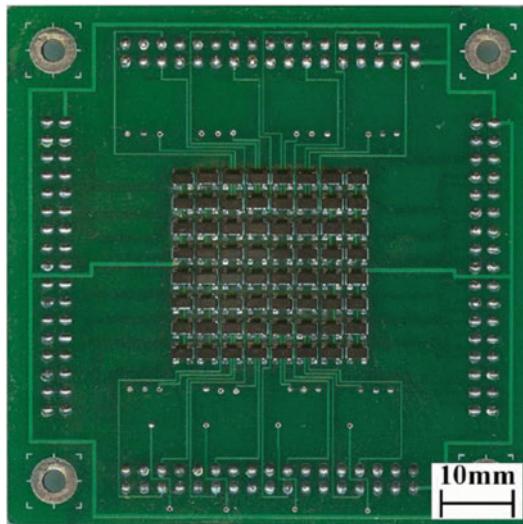
**Fig. 26** Signal from a Hall sensor when a steel wire rope with a broken strand passes through the wire rope tester incorporating parallel NdFeB permanent magnet segments. (Reproduced from Kaur et al. (2018a) with permission from Springer)

the MFL signal. Figure 26 shows the temporal response of the Hall sensor signal as the wire rope is pulled through the wire rope tester. The amplitude of the signal rises when the defect passes across the Hall sensor. The Hall voltages vary from  $-0.08$  to  $0.17$  V with a peak-to-peak value of  $250$  mV. The SNR is  $4.25$  (SNR  $> 2$  for the minimum detectable flaw for the surrounding environment).

## Crack Detection and Profiling

Large structures such as aircraft, automobiles, manufacturing installations, railways, and power equipment suffer fatigue cracks. Detection of crack and its parameters such as size and shape is important for classifying them for better operation and management of these large safety-critical structures. Several studies have been reported where crack profiling is studied to obtain the position, size, and shape. Okolo and Meydan (2017) presented an MFL testing method with Hall sensors to detect hairline cracks in a carbon steel plate hidden  $4$  mm deep below the surface. Oota et al. (2001) developed a scanning Hall sensor microscope with a small footprint of  $50 \times 50 \mu\text{m}^2$  and successfully detected fatigue cracks in mild steel specimen covered under nonmagnetic paint. Magneto-optic methods have been used to inspect large areas whose output can be a black and white binary image (Fitzpatrick et al. 1993). Jun et al. (2007) developed a  $64$  array Hall sensor matrix to inspect partially magnetized paramagnetic materials like austenitic stainless steel whose magnetic domain saturates by external magnetization of  $\delta$ -ferrite. Figure 27 shows the array used as a magnetic camera. It consisted of a magnetic lens to concentrate the distorted magnetic field from the crack. The bias output was mixed with the alternating Hall voltage because the specimen contains partially magnetized structure. The direct response signal was extracted using a high-pass filter (HPF) and a root-mean-squared (RMS) circuit. The RMS Hall voltages obtained from the Hall sensor matrix of  $3.5$  mm spatial resolution is shown in Fig. 28(a). Two peak values

**Fig. 27** Photograph of a 64 array Hall sensor matrix used as a magnetic camera to image stainless steel plates for cracks. (Reproduced from Jun et al. (2007) with permission from the IEEE)



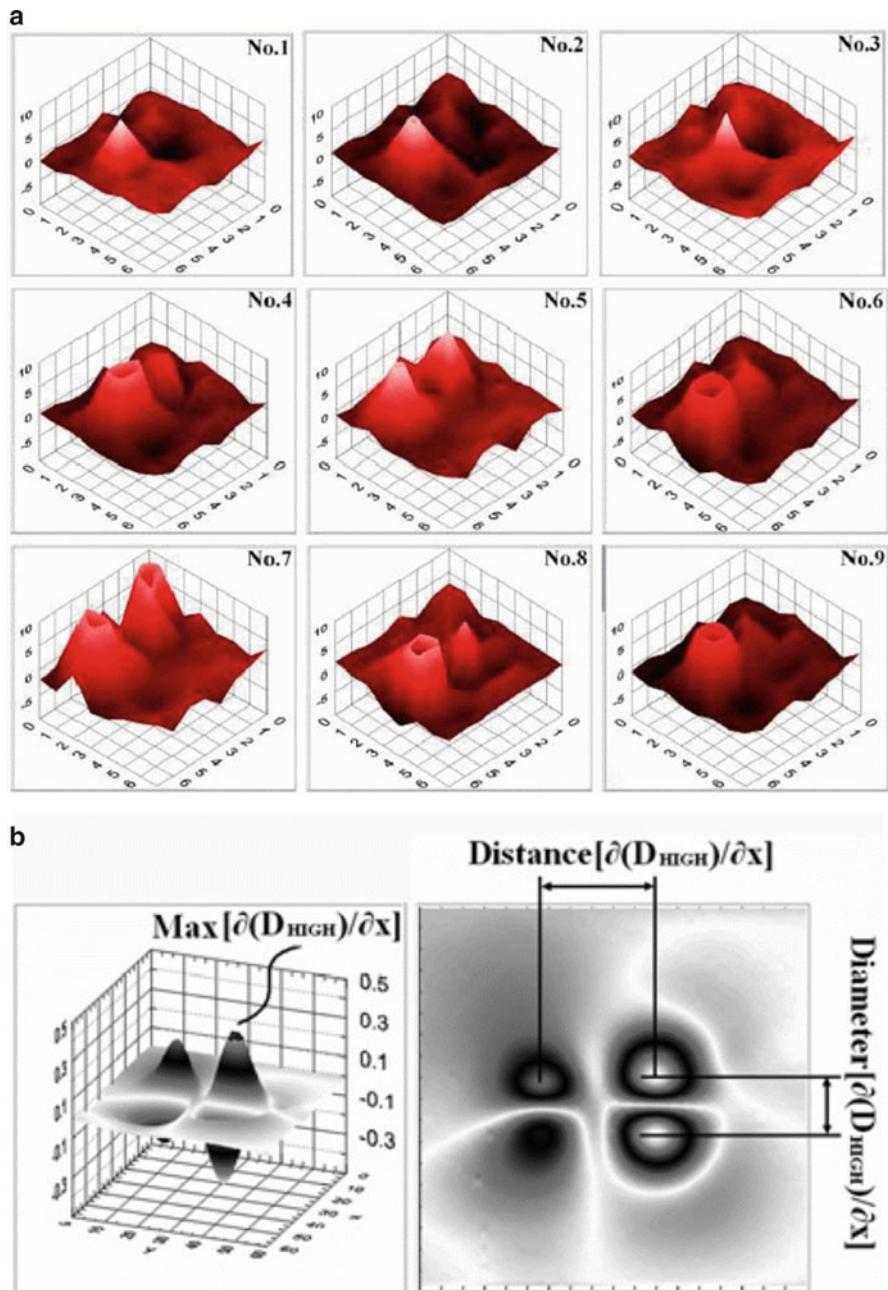
on the scan appeared at both tips of the crack. Jun and Lee (2008) presented a method to determine the maximum and minimum positions of these peaks to locate the crack ends, which indicated the crack length. The located crack is shown in Fig. 28(b) using the identified peaks in the scan. Crack lengths of 5–15 mm and depth 2–9 mm were detected and measured. Simulation models to estimate the crack depth direction have been reported by using the Hall voltages acquired from a magnetic camera (Le et al. 2013).

## Composite and Sandwich Structures

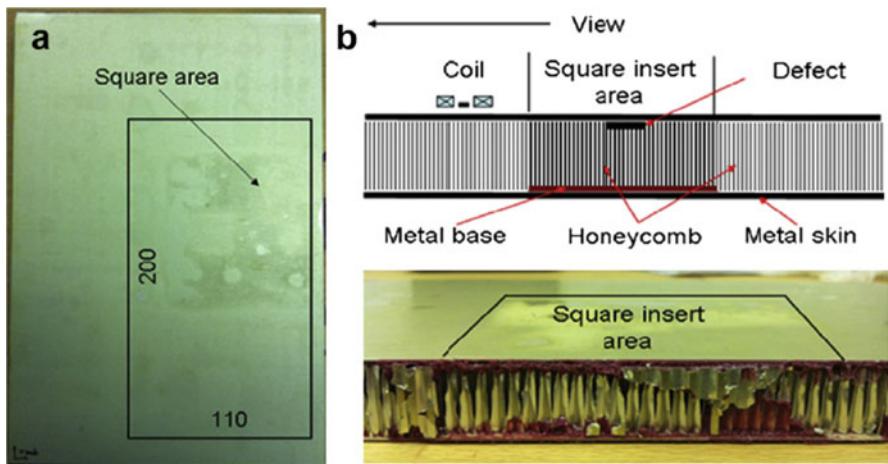
In the last few decades, composite materials have gained immense popularity and replaced metals or alloys in several industries, namely, aviation, automotive, space, and construction owing to their salient properties of light-weight, high-specific stiffness and strength. Despite their high environmental and fatigue resistance, laminated fiber-reinforced polymers (FRP) are often vulnerable to flaws during fabrication and service such as fatigue cracks or disbonds in adhesive metal-composite joints. Hence, there is a need for NDE experts to develop methodologies for inspecting composite materials. Due to low conductivity and material anisotropy, composite materials pose several challenges in eddy current testing. Sandwich structures present additional complexity through attenuation and geometric features. Induction coils measure the derivative of the magnetic field that dies rapidly. Consequently, the deep layer information of composite is buried in the noise.

## Pulsed Eddy Current Excitation and Hall Sensors

The magnetometer measures the magnetic field directly for which He et al. (2014) selected a Hall sensor as the pickup unit to map the magnetic field intensity in the



**Fig. 28** (a) Experimental scan results from an austenitic stainless steel specimen with cracks of different configurations. A Hall sensor matrix was used as a magnetic camera to image the cracks. (Reproduced from Jun et al. (2007) with permission from the IEEE). (b) Location of crack ends in the image determined from the peaks of the Hall voltage distribution. (Reproduced from Jun and Lee (2008) with permission from Springer)



**Fig. 29** Honeycomb sample (a) the backside photo showing the scan area and (b) cross-sectional schematic and photo showing the honeycomb construction, the scan direction, and the location of the defect. (Reproduced from He et al. (2014) with permission from Elsevier)

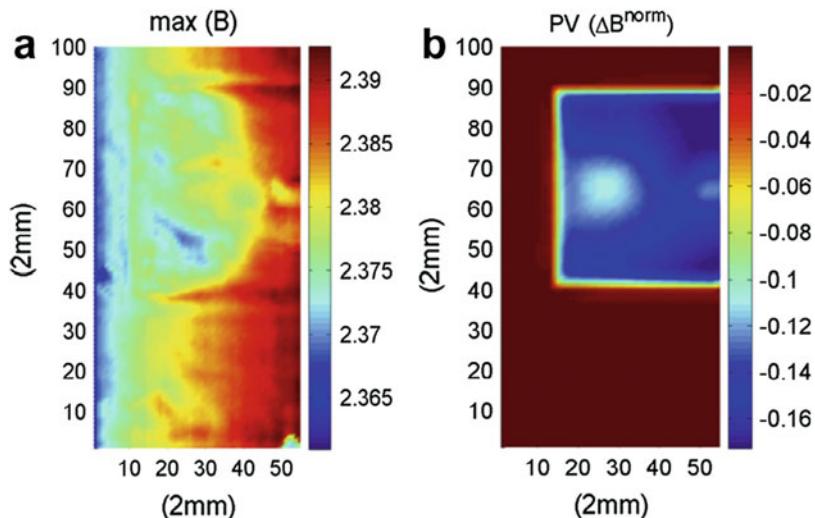
carbon-fiber-reinforced polymer (CFRP) laminates and honeycomb core sandwich panels. Figure 29 shows the specimen and insert type of defect in a honeycomb sandwich panel. In their experimental setup, there is a positive monotonic relationship between peak  $B$  and permeability, due to which conductivity, lift-off, and other parameters easily affect the response. To minimize these effects, the normalized response  $\Delta B_N$  was estimated as

$$\Delta B_N = B / \max(B) - B_R / \max(B_R), \quad (23)$$

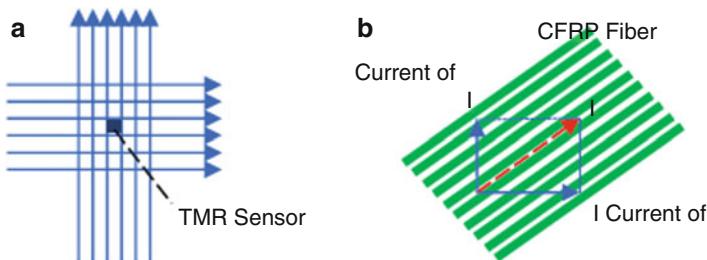
where  $B_R$  is the reference signal. The pulsed eddy current scan using Hall sensors showed the defect locations as shown in Fig. 30. The scan result obtained by  $\Delta B_N$  were clearer because  $\Delta B_N$  was low in a square area including a metal base with higher conductivity, whereas the nonconductive defects decrease  $\Delta B_N$  resulting in the contrast.

### Tunneling Magnetoresistive (TMR) Sensor with Two Orthogonal Excitation Coils

Traditional NDE techniques, such as ultrasonic or X-ray, have to deal with anisotropy of composite materials and the need for coupling into the sample, which limits their applications. Rosell et al. (2018) investigated an eddy current probe, with a high sensitivity TMR sensor located at the center of the two orthogonal coils, for detection of defects in CFRP samples (see Fig. 31). The two planar coils are exciting currents in phase but with different amplitudes, which generates eddy currents in the composite surface layers in any desired direction. The TMR sensor measures the  $z$  component of the induced magnetic field indicating the presence of defects in the sample. Low power requirement and high magnetic field sensitivity of TMR



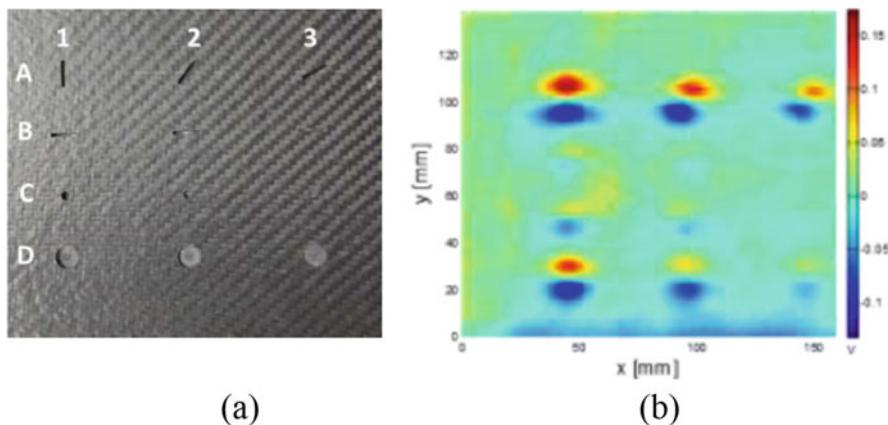
**Fig. 30** Scan results of the backside of honeycomb sample using Hall sensor and pulsed eddy current (a)  $\text{max}(B)$  and (b)  $\Delta B_N$ . (Reproduced from He et al. (2014) with permission from Elsevier)



**Fig. 31** (a) Schematic of EC probe with orthogonal coils and TMR sensor located at the center of the coils. (b) Orthogonal excitation coils to generate currents along controlled directions. (Cited from Rosell et al. (2018))

sensors at room temperature make them promising for NDE applications with fine spatial resolution and high sensitivity (Ye et al. 2018a).

Imaging results from TMR sensor on a 3.2 mm thick, 16-layered [0/90] CFRP sample are shown in Fig. 32. Several rectangular notches of varying angles and flat bottom holes of varying depths were machined in the sample. A sinusoidal current with a frequency of 100 kHz was excited using an Agilent 35500B function generator with peak-to-peak amplitude of 10 V. Use of orthogonal linear excitation coils ensured that the eddy currents could be induced in the sample without rotating the probe mechanically, thereby picking up defect indications in different orientations based on the direction of exciting current. Other magnetic sensors used for



**Fig. 32** (a) CFRP sample with machined defects. (b) Image from TMR sensor data with excitation current oriented at  $90^\circ$ . (Cited from Rosell et al. (2018))

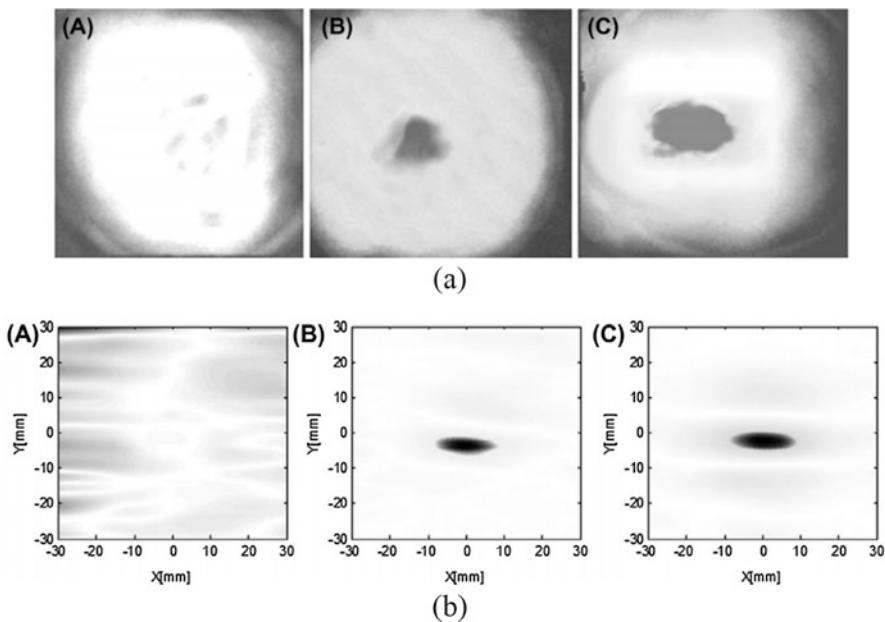
composite inspection have been presented in the literature, such as high-frequency eddy current (HFEC) sensors, which were investigated for permittivity measurement on insulating materials (Gabler et al. 2015). Wikswo et al. (1993) showed that a high-resolution, sensitive SQUID susceptibility imaging system could also be used to measure the susceptibility distribution in a non-conducting material for detection and sizing of flaws.

### SQUID Magnetometer

Wang et al. (2017) recently presented the capability of SQUID magnetometers for NDE of 3-D braided composites by analyzing electrical properties of the materials. It is evident from Fig. 33 that SQUID sensors provide much higher resolution images for impact defects compared to ultrasonic testing which enables accurate characterization of shape and size of flaws. SQUIDs have also been useful for the detection of very small ( $\mu\text{m}$  size) metallic contaminants in either food, water, or industrial products, as presented by Tanaka et al. (2004, 2009). Besides all these applications, the major part of current research in SQUID systems consists of developing long-term but cost-effective cooling mechanisms at its operation site (Hato and Tanabe 2018).

### Multilayer Joints

Aircraft lap joints can contain subsurface defects at the joint interface. The detection schemes must possess the capability to detect surface as well as subsurface defects. Sophian et al. (2002) proposed a pulsed eddy current sensor to detect defects in aircraft lap joints made of aluminum. The wider frequency band of pulsed excitation allowed the detection of deeply hidden subsurface defects. They considered inductive flat coils to generate varying magnetic fields and Hall sensors as pickup sensors.



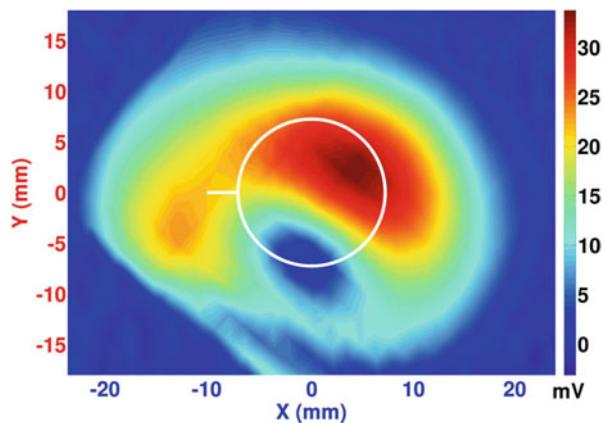
**Fig. 33** (a) Ultrasonic images of 5 mm thick CFRP samples impacted by (A) 2 J, (B) 5 J, and (C) 12 J. (b) Flux image (sensed by SQUID) of the same samples. (Reproduced from Wang et al. (2017))

Hall sensors were selected since they have a much wider magnetic field range. They determined the discrete Fourier transforms of the time series response. Results indicated the peak magnitude from time response around 10 kHz showing peaks corresponding to the defects on the lap joint. The frequency component magnitude indicated the capability to detect subsurface defects.

Aging aircraft fleets face major challenges to ascertain the health of structures. Corrosion under fasteners, cracks from rivet hole, and hole-enlargement can significantly weaken the structure. Eddy current-based MOI technique provides high accuracy, reliability, and easy use to detect deeper defects, cracks, and corrosion. Zeng et al. (2006) presented the determination of subsurface fatigue cracks at rivet sites by using the 3-D finite element simulation model employing Maxwell's equations. From the binary images, a skewness parameter was established to estimate the size of the defect. Observing the advantages of MOI technique, a probability-of-detection study was further done using the skewness function (Deng and Liu 2006; Deng et al. 2012). An enhanced MOI technique, utilizing a laser was further presented with optimized magnetic optic sensor, excitation device, and image processing approach (Cheng et al. 2013).

EC-GMR array probes provide the orientation of defects under fasteners by measuring the normal component of the magnetic field due to defects using an array of GMR sensors (Rifai et al. 2016; Yang et al. 2012). Yang et al. (2015) used rotating field EC-GMR to detect cracks at fasteners in a layered metal joint.

**Fig. 34** C scan of 3-layered aluminum alloy with a notch in the bottom of the middle layer obtained using two coils based rotating field probe and a GMR sensor. (Reproduced from Yang et al. (2015) with permission from the IEEE)



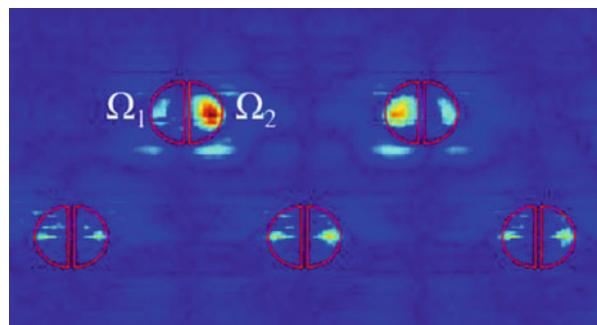
Here, two unidirectional coils oriented in orthogonal directions and carrying currents with  $90^\circ$  phase shift generated the rotating magnetic fields. The magnetic field is in the tangential plane with uniform sensitivity to cracks oriented in an arbitrary direction. Such excitation can be represented as a complex function given by  $J_R = \hat{y}J_0 + j\hat{x}J_0$ . The GMR sensor placed at the center of the probe measured the normal component of the rotating magnetic field given by  $\tilde{B}_R = \hat{z}(\tilde{B}_1 + \tilde{B}_2)$  where  $\tilde{B}_1$  and  $\tilde{B}_2$  are complex valued normal components of the magnetic flux density due to real and complex components of the excitation currents. The system has low-cost, low power requirement with linear sensitivity over a broad range (DC to 1 MHz). A scan of 3-layered aluminum 7075-T5 plate specimen with a slot around the rivet shank in the bottom of the middle layer is shown in Fig. 34. The asymmetry indicating the defect can be seen along the  $x$ -axis at a  $180^\circ$  orientation.

Karpenko et al. (2017) reported a signal enhancement by dual frequency fusion of the scan data to suppress the strong fastener response and improve detection. Ye et al. (2018b) used high-density and large-scale TMR sensor array for magnetic field imaging to detect notches under fastener heads at a distance of 8 mm from the top surface of a multilayer riveted aluminum structure. The TMR array consists of 180 sensing elements with a distance of 0.5 mm between elements and a total length of 90 mm to image a rectangular wide area resulting in a fast inspection speed of 600 mm/min. The data obtained at 1000 Hz was transformed and subtracted from the low-frequency 100 Hz counterpart. Figure 35 confirms the presence of the defect due to asymmetry along the  $x$ -axis.

## Mechanical Properties and Stress

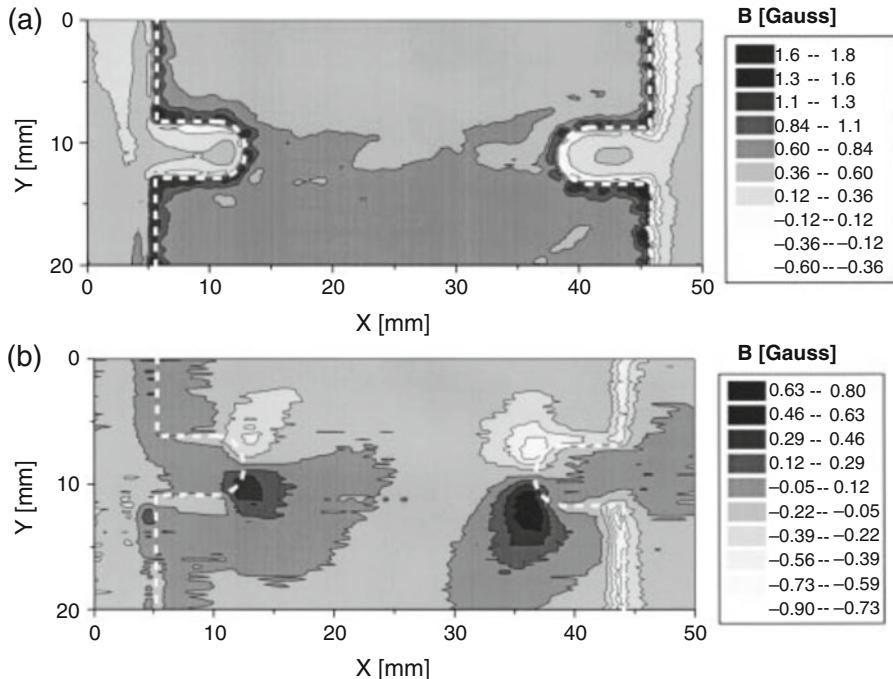
Demands of higher reliability and stability of structural materials have led to the development of NDE techniques to investigate material changes under fatigue, strain, and aging. Plastic deformation arises due to external loading during service. It distorts the crystalline lattice structure permanently creating dislocations by slip

**Fig. 35** C scan of riveted 2-layered sample with the notch in the bottom of the lower layer obtained using a TMR array sensor.  
(Reproduced from Ye et al. (2018b) with permission from IEEE)



process. The plastic deformation degrades the local material properties and its progression may cause failure. Recent progress on the development of SQUID and micro-Hall sensors provides promising techniques over conventional ultrasound and X-ray imaging. Kasai et al. (1997) proposed the detection of Luders bands in mild steels subjected to a strain process using SQUID. Oota et al. (2001) used a micro-Hall sensor with  $50 \times 50 \mu\text{m}^2$  footprint to obtain two-dimensional magnetic images of 304 stainless steel subjected to straining or fatigue process. The plastic deformation was induced in stainless steel plates with a strain rate of 0.001/s along the length of the specimen. Fatigue tests with  $28 \text{ kg/mm}^2$  and frequency of 29.2 Hz were conducted to induce fatigue micro cracks. Hall sensors used in this work were AREPOC: HHP-VP consisting of epitaxial GaAs films with a sensitivity of 3.39 mV/G at a Hall current of 10 mA. The scan step was 0.05 mm with fixed lift-off distance of 0.5 mm. The images of an unstrained plate show a clear boundary of the specimen with the shape of the notches as shown in Fig. 36. With the application of plastic strain, a significant change in the images with distributions involving two peaks is seen. This is because the 304 stainless steel transforms from the paramagnetic austenite phase to the ferromagnetic martensite phase when subjected to plastic strain at room temperature. This phase also called as strain-induced martensitic transformation was confirmed from microstructure studies.

Sharatchandra Singh et al. (2018) also demonstrated the measurement of deformation-induced magnetic fields to detect plastic deformation with SNR greater than 20 dB using GMR sensors. They used a specimen made of S235JR carbon steel and a spin-valve GMR sensor array connected in a gradiometric configuration. The scan was performed with a step size of 1  $\mu\text{m}$  to obtain self-magnetic leakage field (SMLF) images of the tensile specimens after tensile loads of 0 kN and 13.1 kN. Figure 37 (a) and (b) shows the self-magnetic leakage field map of the specimen with 0 and 13.1 kN, respectively. Randomly distributed small magnetic stray fields of the order of  $H = \pm 20 \text{ A/m}$  tensile load (0 kN) is seen in the unloaded specimen. The specimens start exhibiting deformation-induced magnetization on the surface of the specimens with 13.1 kN load, which is a work hardening region. Specimens undergo larger plastic deformation in the reduced cross section. The effective magnetic field induced  $H_s$  due to stress  $S$  is given by (Dong et al. 2009)

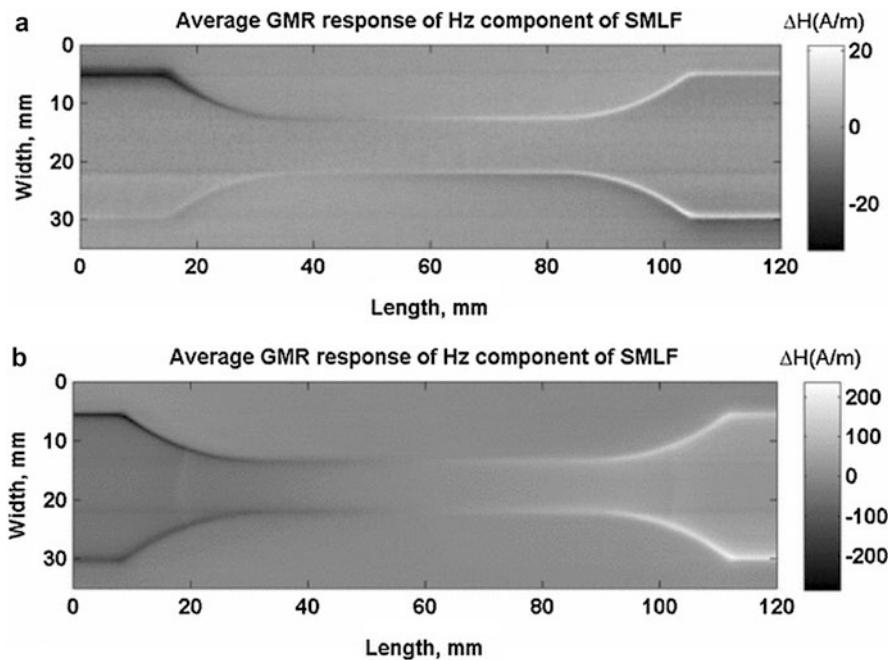


**Fig. 36** Magnetic images of spontaneous magnetization on a surface of stainless steel plate A subjected to a plastic strain obtained using a Hall sensor (a) unstrained; (b) 2.5% strained. (Reproduced from Oota et al. (2001) with permission from Springer)

$$H_S = \frac{3S}{2\mu_0} d \frac{\partial \lambda}{\partial M}, \quad (24)$$

where  $\lambda$  is magnetostrictive coefficient and  $M$  is the magnetization. Stress turns the magnetic domains toward its direction creating self-induced magnetic signals with opposite signs at the two ends of the specimen. The effective magnetic permeability decreases due to the occurrence of dislocations. Dislocations act as pinning sites for movement of domain walls leading to higher MF leakage.

Reports are available suggesting 80% of accidents result from local stress concentration (Liu et al. 2015b). The reason is attributed to the reduction of strength, fatigue life, dimensional stability, and corrosion resistance. Thus, it is important to detect, monitor, and evaluate the initiation and propagation of stress to predict and prolong service life. Compared to other techniques, pulsed ECT has good linearity and high precision with the capability to measure stress (Morozov et al. 2010). Stress concentration can lead to variations in conductivity and permeability. According to the piezoresistive effect, variations in conductivity are low (around 5%) (Morozov et al. 2009). The stress mainly affects the magnetic permeability (Prokopovych and Osadchuk 2011).



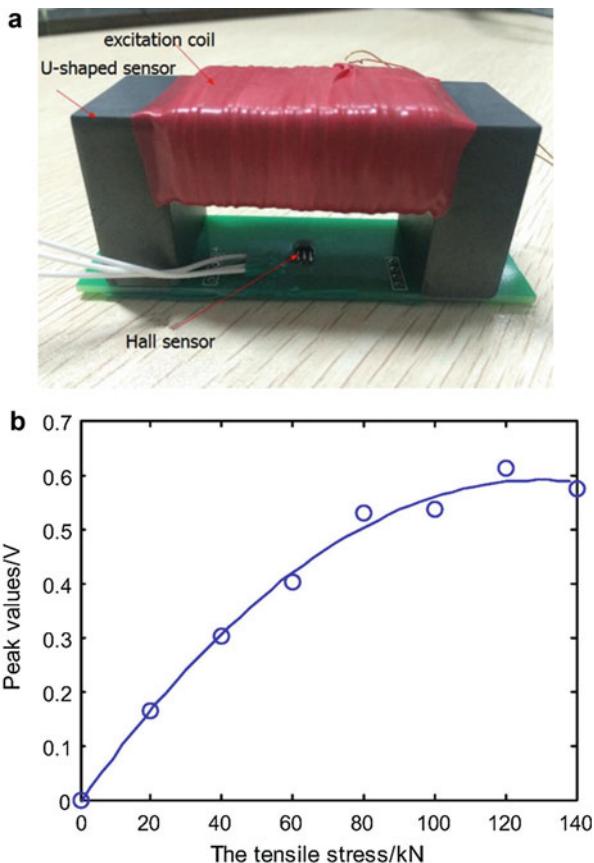
**Fig. 37** Self-magnetic leakage field (SMLF) images of the tensile specimens obtained using GMR sensors after tensile loads of (a) 0 kN and (b) 13.1 kN. (Reproduced from Sharatchandra Singh et al. (2018) with permission from Springer)

Zhou et al. (2017) proposed a U-shaped sensor to produce a large and uniform excitation field between the two poles of the yoke. They studied the effect of magnetic permeability and conductivity on the eddy current response. Figure 38(a) shows the U-shaped sensor with a U-shaped magnetic core, coil, and a Hall sensor YS96B placed in the middle of the U-shaped core. The Hall sensor was supplied with 5 V and its output was filtered and amplified. The peak values of the response signal variation with loads of 0–140 kN are shown in Fig. 38(b). The signals were obtained with excitation at 5 Hz. Due to interference in the external magnetic field, error in the results was seen. The increase in tensile stress increased the peak value with an approximately linear relationship in the lower range from 0 to 100 kN. Beyond 100 kN, the peak value displayed the yield effect due to plastic deformation. Similar detection of residual stress due to shot peening in S235JR carbon steel specimens was reported by Sharatchandra Singh et al. (2018) using GMR sensors.

## Rail Wheel Defects

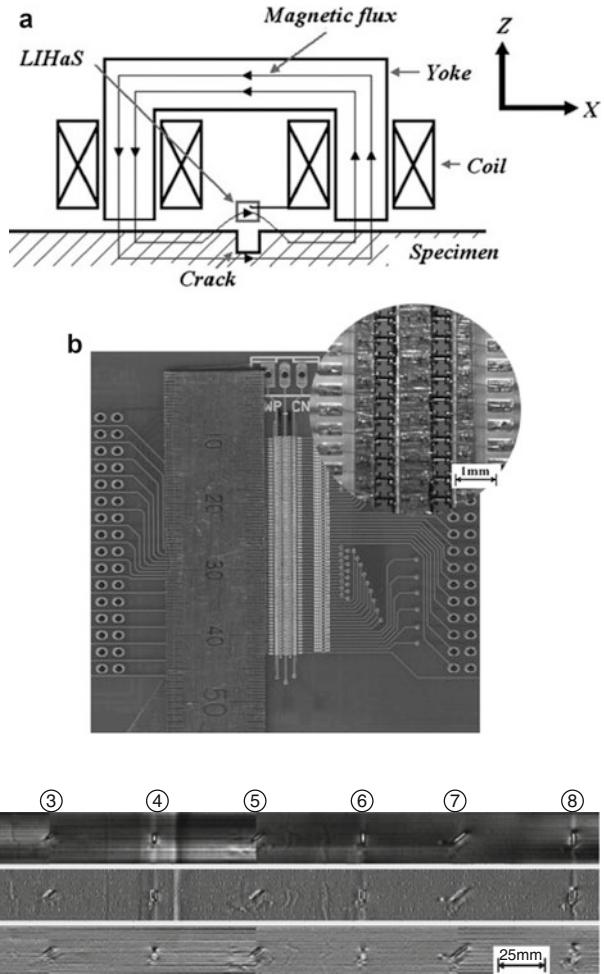
Rail wheels suffer damage such as fatigue cracks, internal defects, crushing, side grinding, corrosion, stripping, transverse defects in the thread, and cracks on the flange

**Fig. 38** (a) The U-shaped sensor comprising the U-shaped coil and Hall sensor to detect stress concentration in the magnetic material. (b) The variation of the peak value of the Hall sensor with tensile stresses at an excitation frequency of 5 Hz.  
(Reproduced from Zhou et al. (2017) with permission from Elsevier)



and chamfer. Compared to other techniques, eddy current-based detection schemes meet the requirement of the railway industry, which specifies just 1 h for testing. Hwang et al. (2009) proposed a differential Hall sensor array-based scan type of magnetic camera to satisfy this need. To increase the detection ability, they used a yoke-type magnetizer as a magnetic source. To concentrate the flux around the crack, they used a magnetic lens (see Fig. 39(a)). Figure 39(b) shows a wafer-supported, two linearly integrated Hall sensor arrays (LIHaS) used to achieve high-speed testing. A total of 64 InSb Hall sensors with an interval of 0.52 mm were arrayed on a NiZn ferrite wafer to give a high resolution. The  $\partial V_H / \partial x$ ,  $\partial^2 V_H / \partial x^2$ , and  $\partial^2 V_H / \partial x \partial y$  value was obtained directly by connecting two  $x$ -directional sensors by wire ball bonding. The output amplified 911 times had a range of  $\pm 4$  mT with 2.51 V/mT sensitivity. Figure 40 shows the measurement images from which the crack locations can be visualized. The horizontal lines were minimized in the image corresponding to the differential quantities. Using these images, they further correlated the crack width and length to the observed values  $\partial V_H / \partial x$  from the images. The volume of the crack was estimated by integrating the minimum and maximum values as

**Fig. 39** (a) Two parallel arrayed differential type linearly integrated 64 InSb Hall sensor array on a NiZn ferrite wafer. (b) A small, yoke-type, electromagnetic coil operated by an alternating current. (Reproduced from Hwang et al. (2009) with permission from Elsevier)



**Fig. 40** Experimental wheel results from using the linearly integrated 64 InSb Hall sensor array with pulse generation at 0.43 mm. (Reproduced from Hwang et al. (2009) with permission from Elsevier)

$$V_C = C_1 \left[ \sum_{i=1}^{L_c/S} \left( \text{Max} \left( \frac{\partial V_{H,i}}{\partial x} \right) + \text{Abs} \left( \text{Min} \left( \frac{\partial V_{H,i}}{\partial x} \right) \right) \right) \right]^{C_2}, \quad (25)$$

where  $L_c$  is the crack length,  $S$  is a scanner parameter, and  $C_1$  and  $C_2$  are constants depending on the material, lift-off, Hall coefficient, and magnetizer intensity. Several works on the classification of the cracks based on the location, inclination, width, depth, and length have been reported. The classification is done using the  $x$ ,  $y$ , and  $z$  components in the magnetic leakage field (Wang et al. 2014). The challenge lies in the effective measurement of 3-D data and full utilization of the data. Chen et al.

(2018) have presented a rail crack recognition based on the decision from an adaptive weighting multi-classifier fusion algorithm. They used Hall sensor arrays to obtain the signals of different directions. First, a support vector machine is used to classify MFL signals from single-direction and single-channel. Adaptive weighting is assigned by finding the entropy calculated by the posterior probability of different support vector machines. By fusing classification results of different channels and directions, a weighted majority vote strategy helps in the decision-making. The method required fewer training samples and fewer channels of the signals.

## Metrology

Eddy current methods have been employed to measure the thickness of conducting materials to detect wall thinning due to corrosion (Pinotti and Puppin 2014; Cheng 2014). The slope of the curve in the impedance plane is a phase quantity, which is linearly related to the thickness of the target when standard penetration depth is achieved. Pulsed eddy current uses broadband excitation, which provides more information when signals are interpreted in the time-frequency domain. Coatings have been successfully characterized for their thickness and conductivity (Tai et al. 1996; Zhang et al. 2015). Peak values of pulsed eddy currents are affected by lift-off and depth variations (Yu et al. 2014). The phase signature of the coil is immune to lift-off and outperforms the voltage spectra with significant information related to defect and density (Bai et al. 2014; Zhao et al. 2011). Fan et al. (2017) used a Hall sensor rather than a pickup coil to eliminate the effect of lift-off (see Fig. 41(a)). The theoretical model relates the phase to sample thickness through the reflection coefficient as

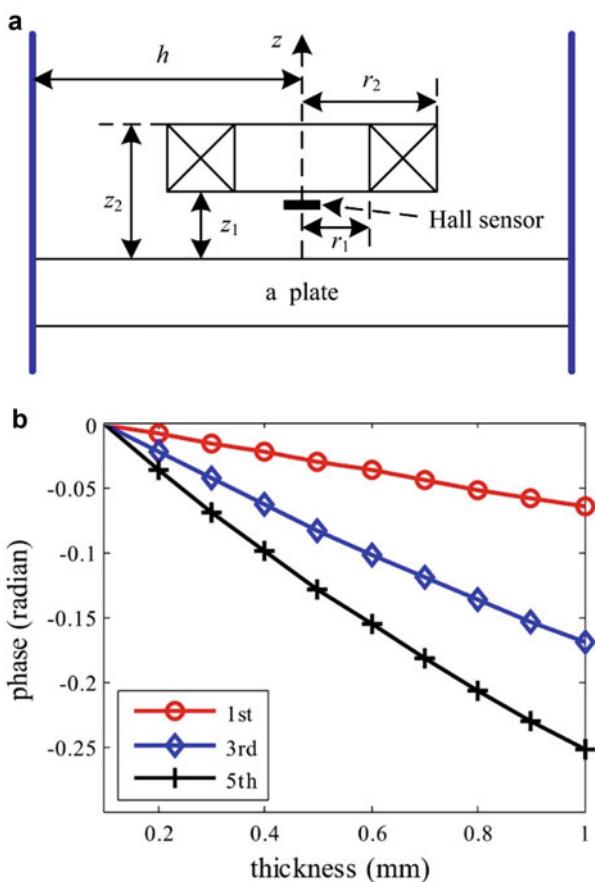
$$R'_{0i,1i} = \frac{(\lambda_{0i}\mu_r)^2 - \lambda_{1i}^2 + e^{-2\lambda_{1i}d} [\lambda_{1i}^2 - (\lambda_{0i}\mu_r)^2]}{(\lambda_{0i} + \lambda_{0i}\mu_r)^2 - e^{-2\lambda_{1i}d}(\lambda_{0i} - \lambda_{0i}\mu_r)^2}, \quad (26)$$

where  $\lambda_{1i} = \sqrt{\lambda_{0i}^2 + j\omega\sigma\mu}$ ,  $\lambda_{0i}$  is the root,  $\mu_r$  is the relative permeability,  $\sigma$  is the electric conductivity,  $\mu$  is the magnetic permeability, and  $d$  is the thickness of the plate. To bring out the effect of thickness, the difference of pulsed eddy current signal was estimated as  $\Delta B(t) = B(t) - B_{air}(t)$ . The phase of the spectral difference signal for different sample thicknesses was calculated as  $\mathcal{O}_{PEC}(n) = \text{Im}[fft(\Delta B)]/\text{Re}[fft(\Delta B)]$ , where  $n$  denotes the index of the frequency component. Fig. 41(b) shows the sensitivity of phase to the sample thickness for first, second, and third frequency components.

## Clinical Applications

High-frequency EM methods are mostly used as a standard practice in medical applications for noninvasive imaging of internal organs such as the X-ray or microwave systems, due to their obvious advantage of higher penetration capabilities through skin and tissues. Among the low-field NDE sensors, SQUIDs have been found suitable for medical applications. A small pickup coil of diameter around

**Fig. 41** (a) Schematic of a pulsed eddy current air core coil over a plate and a Hall sensor placed within the coil. (b) Phase of pulsed eddy current spectral components due to sample thickness, estimated from the Hall sensor response. (Reproduced from Fan et al. (2017) with permission from Elsevier)



3 mm, extremely low coil-to-sample distance made it a popular technique in medical systems compared to its counterpart low-field methodologies. In the past decade, SQUIDS have shown promise in several clinical applications such as magnetoencephalography (MEG) which is a passive, noninvasive technique for mapping of magnetic fields generated by neural activity in human brains (Xie et al. 2015). The high sensitivity of SQUID magnetometers enables sensing of a very weak magnetic field produced by currents flowing through myocardial fibers during cardiac activities (Tavarozzi et al. 2002). Besides, Merwa et al. (2004) presented magnetic induction tomography (MIT) as a method to detect brain edema through a feasibility study.

## Summary

Several types of low-field magnetic sensors have been presented for a wide range of NDE applications, each of them having their own advantages over the others. On the one hand, SQUID sensors are capable of measuring very low magnetic field and

hence are well suited for biomagnetic applications. However, due to the requirement of a cooling environment, these sensors can be expensive especially if a coolant other than liquid nitrogen is required to initiate superconductivity in the materials. On the other hand, fluxgate sensors, which are less expensive than SQUIDs are used in military applications requiring small magnetic field measurements. MR sensors and Hall sensors are developed to measure fields of medium and high strength, respectively. Induction sensors have the widest range compared to any other sensor and can be tailored toward any application, with obvious limitations on the sensitivity and dimensions requirements of the sensors. The GMR sensors are most promising in terms of sensitivity and cost. The TMR sensors are in the stage of development, and efforts are needed to overcome the handling abilities by improving the tolerance to higher voltage levels without shorting.

**Acknowledgment** This work was partially supported by the US Department of Energy under the award number: DE-FE0031650.

---

## References

- Afzal M, Udupa S (2002) Advanced signal processing of magnetic flux leakage data obtained from seamless gas pipeline. *NDT E Int* 35:449–457
- Aguila-Muñoz J, Espina-Hernández JH, Pérez-Benítez JA, Caleyo F, Hallen JM (2016) A magnetic perturbation GMR-based probe for the nondestructive evaluation of surface cracks in ferromagnetic steels. *NDT&E Int* 79:132–141
- Arismendi NOR, Pacheco ER, López OP, Espina Hernández JH, Benitez JAP (2018) Classification of artificial near-side cracks in aluminium plates using a GMR-based eddy current probe. In: 2018 international conference on electronics, communications and computers (CONIELECOMP), Cholula, Mexico
- Bai X, Fang Y, Lin W, Wang L, Ju BF (2014) Saliency-based defect detection in industrial images by using phase spectrum. *IEEE Trans Ind Inform* 10:2135–2145
- Bailey J, Long N, Hunze A (2017) Eddy current testing with giant magnetoresistance (GMR) sensors and a pipe-encircling excitation for evaluation of corrosion under insulation. *Sensors* 17:2229. <https://doi.org/10.3390/s17102229>
- Behun L, Smetana M, Capova K (2018) Comparison of detection abilities between fluxgate and GMR magnetometer in inverse ECT of deep lying cracks. In: 2018 ELEKTRO, Mikulov
- Birosca S, Buffiere JY, Garcia-Pastor FA, Karadge M, Babout L, Preuss M (2009) Three-dimensional characterization of fatigue cracks in Ti-6246 using X-ray tomography and electron backscatter diffraction. *Acta Mater* 57:5834–5847
- Buchanan DS, Crum DB, Cox D, Wikswo JP (1989) MicroSQUID: a close-spaced four channel magnetometer. In: Advances in biomagnetism. Springer, Boston, pp 677–679
- Campbell WH (1969) Induction loop antennas for geomagnetic field variation measurements. ESSA technical report ERL123-ESL6
- Carreon H (2006) Thermoelectric detection of the magnetic field by fluxgate gradiometer on subsurface tin inclusions embedded in a copper bar. *NDT&E Int* 39:22–28
- Carreon H, Nagy PB, Nayfeh AH (2000) Thermoelectric detection of spherical tin inclusions in copper by magnetic sensing. *J Appl Phys* 88:6495
- Caruso MJ, Bratland T, Smith CH, Schneider R (1998) A new perspective on magnetic field sensing. *Sensors-Peterborough* 15:34–47
- Chen W, Liu W, Li K, Wang P, Zhu H, Zhang Y, Hang C (2018) Rail crack recognition based on adaptive weighting multi-classifier fusion decision. *Measurement* 123:102–114

- Cheng WY (2014) Pulsed eddy current characterization of local wall-thinning. *Int J Appl Electromagn Mech* 45:179–184
- Cheng Y, Deng Y, Bai L, Chen K (2013) Enhanced laser-based magneto-optic imaging system for nondestructive evaluation applications. *IEEE Trans Instrum Meas* 62:1192–1198
- Clarke DB (1999) Demagnetization factors of ringcores. *IEEE Trans Magn* 35:4440–4444
- Coramik M, Ege Y (2017) Discontinuity inspection in pipelines: a comparison review. *Measurement* 111:359–373
- Deng Y, Liu X (2011) Electromagnetic imaging methods for nondestructive evaluation applications. *Sensors* 11:11774–11808
- Deng Y, Liu X, Fan Y, Zeng Z, Udupa L, Shih W (2006) Characterization of magneto-optic imaging data for aircraft inspection. *IEEE Trans Magn* 42:3228–3230
- Deng Y, Liu X, Udupa L (2012) Magneto-optic imaging for aircraft skins inspection: a probability of detection study of simulated and experimental image data. *IEEE Trans Reliab* 61:901–908
- Dogaru T, Smith ST (2001) Giant magnetoresistance-based eddy-current sensor. *IEEE Trans Magn* 37:3831–3838
- Dong L, Xu B, Dong S, Song L, Chen Q, Wang D (2009) Stress dependence of the spontaneous stray field signals of ferromagnetic steels fields. *NDT&E Int* 42:323–327
- Ege Y, Coramik M, Kabadayi M, Citak H, Kalender O, Yürüklü E, Kurt U, Nazlibilek S (2016) Anomaly detection with low magnetic flux: a fluxgate sensor network application. *Measurement* 81:43–56
- Espinosa AG, Rosero A, Cusido J, Romeral L, Ortega JA (2007) Fault detection by means of Hilbert Huang transform of the stator current in a PMSM with demagnetization. *IEEE Trans Energy Convers* 25:312–318
- Fan M, Cao B, Sunny AI, Li W, Tian G, Ye B (2017) Pulsed eddy current thickness measurement using phase features immune to liftoff effect. *NDT&E Int* 86:123–131
- Fitzpatrick G, Thome D, Skaugset R, Shih E, Shih W (1993) Magneto-optical/Eddy current imaging of aging aircraft – a new NDI technique. *Mater Eval* 12:1402–1407
- Fitzpatrick GL, Thome DK, Skaugset RL, Shih WCL (1996) Magneto-Optic/Eddy Current Imaging of Subsurface Corrosion and Fatigue Cracks in Aging Aircraft, Review of Progress In Quantitative Nondestructive Evaluation, Plenum Press, New York, vol 15, pp 1159–1166
- Gabler S, Heurer H, Heinrich G (2015) Measuring and imaging permittivity of insulators using high-frequency eddy-current devices. *IEEE Trans Instrum Meas* 64:2227–2238
- García-Martín J, Gómez-Gil J, Vázquez-Sánchez E (2011) Non-destructive techniques based on eddy current testing. *Sensors* 11:2525–2565
- Goktas T, Lee KW, Zafarani M, Akin B (2016) Analysis of magnet defect faults in permanent magnet synchronous motors through fluxgate sensors. In 2016 IEEE applied power electronics conference and exposition (APEC), Long Beach, 20–24
- Han JS, Park JH (2004) Detection of corrosion steel under an organic coating by infrared photograpy. *Corros Sci* 46:787–793
- Hato T, Tanabe K (2018) Long period of time cooling technique using liquid nitrogen for HTS-SQUID system. *IEEE Trans Appl Supercond* 28:1–4
- Hatsukade Y, Tanaka S (2016) Mobile NDE system utilizing robust HTS-SQUID magnetometer for use in unshielded environments. *IEEE Trans Appl Supercond* 26:1–4
- Hatsukade Y, Kobayashi T, Nakae S, Masutani N, Tanaka Y (2017) Novel remote NDE technique for pipes combining HTS-SQUID and ultrasonic guided wave. *IEEE Trans Appl Supercond* 27:1–4
- He Y, Tian G, Pan M, Chen D (2014) Non-destructive testing of low-energy impact in CFRP laminates and interior defects in honeycomb sandwich using scanning pulsed eddy current. *Compos Part B Eng* 59:196–203
- Hwang J, Lee J, Kwon S (2009) The application of a differential-type Hall sensors array to the nondestructive testing of express train wheels. *NDT & E Int* 42:34–41
- Jander A, Smith C, Schneider R, (2005) Magnetoresistive sensors for nondestructive evaluation. In: Proceedings of SPIE 5770, Advanced sensor technologies for nondestructive evaluation and structural health monitoring, San Diego

- Jenks WG, Sadeghi SSH, Wikswo JP (1997) SQUIDs for nondestructive evaluation. *J Phys D Appl Phys* 30:293
- Jun J, Lee J (2008) Nondestructive evaluation of a crack on austenitic stainless steel using a sheet type induced current and a Hall sensor array. *J Mech Sci Technol* 22:1684–1691
- Jun J, Hwang J, Lee J (2007) Quantitative nondestructive evaluation of the crack on the austenite stainless steel using the induced eddy current and the Hall sensor array. In: 2007 IEEE instrumentation & measurement technology conference IMTC, Warsaw
- Kapsalis VC (2017) Advances in magnetic field sensors. *IOP Conf Series* 939:012026
- Karpenko O, Ye C, Udpal L (2017) Dual frequency fusion for defect signal enhancement in EC-GMR inspection of riveted multilayer structures. *NDT & E Int* 92:97–103
- Kasai N, Isikawa N, Yamakawa H, Chinose K, Nakayama S, Odagawa A (1997) Nondestructive detection of dislocations in steel using a SQUID gradiometer. *IEEE Trans Appl Supercond* 7:2315–2318
- Kaur A, Gupta A, Aggarwal H, Arora K, Garg N, Sharma M, Sharma S, Aggarwal N, Sapra G, Goswamy JK (2018a) Non-destructive evaluation and development of a new wire rope tester using parallelly magnetized NdFeB magnet segments. *J Nondestruct Eval* 37:1–8
- Kaur A, Gupta A, Aggarwal H, Sharma M, Sharma S, Aggarwal N, Sapra G, Goswamy JK (2018b) Selection of a Hall sensor for usage in a wire rope tester. In: Computational signal, processing and analysis. Springer Nature Singapore Pte Ltd, Singapore, pp 361–371
- Kim HM, Heo CG, Cho SH, Park GS (2018a) Determination scheme for accurate defect depth in underground pipeline inspection by using magnetic flux leakage sensors. *IEEE Trans Magn* 54:6202805
- Kim HM, Yoo HR, Park GS (2018b) A new design of MFL sensors for self-driving NDT robot to avoid getting stuck in curved underground pipelines. *IEEE Trans Magn* 54:6202705
- Le M, Lee J, Jun J, Kim J, Moh S, Shin K (2013) Hall sensor array based validation of estimation of crack size in metals using magnetic dipole models. *NDT&E Int* 53:18–25
- Lee SJ, Song SH, Jiles DC, Hauser H (2005) Magnetooptic sensor for remote evaluation of surfaces. *IEEE Trans Magn* 41:2257–2259
- Lenz J, Edelstein S (2006) Magnetic sensors and their applications. *IEEE Sensors J* 6:631–649
- Li W, Yuan X, Chen G, Ge J, Yin X, Li K (2016) High sensitivity rotating alternating current field measurement for arbitrary-angle underwater cracks. *NDT&E Int* 79:123–131
- Li F, Feng J, Zhang H, Liu J, Lu S, Ma D (2018) Quick reconstruction of arbitrary pipeline defect profiles from MFL measurements employing modified harmony search algorithm. *IEEE Trans Instrum Meas* 67:2200–2213
- Liu X, Deng Y, Zeng Z, Udpal L, Udpal SS (2009) Model-based inversion technique using element-free Galerkin method and state space search. *IEEE Trans Magn* 45:1486–1489
- Liu B, Cao Y, Zhang H, Lin YR, Sun WR, Xu B (2015a) Weak magnetic flux leakage: A possible method for studying pipeline defects located either inside or outside the structures. *NDT & E Int* 74:81–86
- Liu B, Fu Y, Jian R (2015b) Modelling and analysis of magnetic memory testing method based on the density functional theory. *Nondestr Test Eval* 30:13–25
- Lu S, Feng J, Li F, Liu J (2017) Precise inversion for the reconstruction of arbitrary defect profiles considering velocity effect in magnetic flux leakage testing. *IEEE Trans Magn* 53:6201012
- Meola C (2007) A new approach for estimation of defects detection with infrared thermography. *Mater Lett* 61:747–750
- Merwa R, Hollaus K, Oszkar B, Scharfetter H (2004) Detection of brain oedema using magnetic induction tomography: a feasibility study of the likely sensitivity and detectability. *Physiol Meas* 25:347–354
- Miri-Hakimabad H, Panjeh H, Vejdanioghreian A (2008) Experimental optimization of a landmine detection facility using PGNA method. *Nucl Sci Tech* 19:109–112
- Morozov M, Tian GY, Edgar D (2009) Comparison of PEC and SFEC NDE techniques. *Nondestr Test Eval* 24:153–164
- Morozov M, Tian GY, Withers PJ (2010) The pulsed eddy current response to applied loading of various aluminium alloys. *NDT & E Int* 43:493–500

- Nagendran R, Mohanty I, Thanikai Arasu V, Baskaran R (2018) Transient eddy current NDE system based on fluxgate sensor for the detection of defects in multilayered conducting material. *J Nondestruct Eval* 37:52
- Nair NV, Melapudi VR, Jimenez HR, Liu X, Deng Y, Zeng Z, Udupa L, Moran TJ, Udupa SS (2006) A GMR-based eddy current system for NDE of aircraft structures. *IEEE Trans Magn* 42:3312–3314
- Okolo CK, Meydan T (2017) Axial magnetic field sensing for pulsed magnetic flux leakage hairline crack detection and quantification. In: 2017 IEEE sensors, Glasgow
- Oota A, Miyake K, Sugiyama D, Aoki H (2001) Observation of changes in magnetic images due to a strain or fatigue progress in austenite stainless steels using a scanning Hall-sensor microscope. *J Nondestruct Eval* 20:87–94
- Orlando L, Slob E (2009) Using multicomponent GPR to monitor cracks in a historical building. *J Appl Geophys* 67:327–334
- Ostafin M, Nogaj B (2007) 14N-NQR based device for detection of explosives in landmines. *Measurement* 40:43–54
- Pelkner M, Reimund V, Erthner T, Panke N, Kreutzbruck M (2014) Automated inspection of surface breaking cracks using GMR sensor arrays. *AIP Conf Proc* 1581:1393–1399
- Piao G, Guo J, Hu T, Leung H, Deng Y (2019) Fast reconstruction of 3-D defect profile from MFL signals using key physics-based parameters and SVM. *NDT & E Int* 103:26–38
- Pinotti E, Puppin E (2014) Simple lock-in technique for thickness measurement of metallic plates. *IEEE Trans Instrum Meas* 63:479–484
- Popovic RS (2003) Hall effect devices. Taylor & Francis Group, LLC, Boca Raton
- Prance RJ, Clark TD, Prance H (2000) Ultra low noise induction magnetometer for variable temperature operation. *Sens Actuators A* 85:361–364
- Prieto MD, Espinosa AG, Ruiz JR, Urresty JC, Ortega JA (2011) Feature extraction of demagnetization faults in permanent-magnet synchronous motors based on box-counting fractal dimension. *IEEE Trans Ind Electron* 58:1594–1605
- Prokopovych IB, Osadchuk VA (2011) Construction of the equation for the influence of stresses on magnetic permeability by the method of free deformation. *Mater Sci* 46:517–524
- Pullen AL, Charlton PC, Pearson NR, Whitehead NJ (2018) Magnetic flux leakage scanning velocities for tank floor inspection. *IEEE Trans Mag* 54:7402608
- Ramos HG, Ribeiro AL (2014) Present and future impact of magnetic sensors in NDE. *Procedia Eng* 86:406–419
- Ramos HMG, Postolache O, Alegría FC, Ribeiro AL (2009) Using the skin effect to estimate cracks depths in metallic structures. In: I2MTC 2009 – international instrumentation and measurement, Singapore
- Rifai D, Abdalla AN, Ali K, Razali R (2016) Giant magnetoresistance sensors: a review on structures and non-destructive eddy current testing applications. *Sensors* 16:298
- Riggs LS, Mooney JE, Lawrence DE (2001) Identification of metallic mine-like objects using low frequency magnetic fields. *IEEE Trans Geosci Remote Sens* 39:56–66
- Riley CD, Jewell GW, Howe D (2000) Design of impulse magnetizing fixtures for the radial homopolar magnetization of isotropic NdFeB ringmagnets. *IEEE Trans Magn* 36:3846–3857
- Ripka P (1992) Review of fluxgate sensors. *Sens Actuators A* 33:129–141
- Ripka P (2003) Advances in fluxgate sensors. *Sens Actuators A* 106:8–14
- Ripka P, Primdahl F (2000) Tuned current-output fluxgate. *Sens Actuators A* 82:160–165
- Rosell A, Ye C, Deng Y, Udupa L, Udupa S (2018) Linear excitation eddy current probe for multi-layered CFRP inspection. In: 12th European conference on non-destructive testing, Gothenburg
- Sasi B, Arjun V, Mukhopadhyay CK, Rao BPC (2018) Enhanced detection of deep seated flaws in 316 stainless steel plates using integrated EC-GMR sensor. *Sens Actuators A* 275:44–50
- Sharatchandra Singh W, Purnachandra Rao B, Mukhopadhyay CK, Jayakumar T (2015) Detection of localized damage in water wall tubes of thermal power plants using GMR sensor array based magnetic flux leakage technique. *J Nondestruct Eval* 34:1–7

- Sharatchandra Singh W, Stegemann R, Kreutzbruck M, Mukhopadhyay CK, Purnachandra Rao B (2018) Mapping of deformation-induced magnetic fields in carbon steels using a GMR sensor based metal magnetic memory technique. *J Nonddestruct Eval* 37:1–8
- Shi Y, Zhang C, Li R, Cai M, Jia G (2015) Theory and application of magnetic flux leakage pipeline detection. *Sensors* 15:31036–31055
- Smith RA, Hugo GR (2001) Transient eddy-current NDE for ageing aircraft. *Insight* 43:14–20
- Smith CH, Schneider RW, Dogaru T, Smith ST (2003) GMR magnetic sensor arrays for NDE eddy current testing. *AIP Conf Proc* 657:419–426
- Sophian A, Tian GY, Taylor D, Rudlin J (2001) Electromagnetic and eddy current NDT: a review. *Insight* 43:302–306
- Sophian A, Tian GY, Taylor D, Rudlin J (2002) Design of a pulsed eddy current sensor for detection of defects in aircraft lap-joints. *Sens Actuators A* 101:92–98
- Suresh V, Abudahir A, Daniel J (2017) Development of magnetic flux leakage measuring system for detection of defect in small diameter steam generator tube. *Measurement* 95:273–279
- Tai CC, Rose JH, Moulder JC (1996) Thickness and conductivity of metallic layers from pulsed eddy-current measurements. *Rev Sci Instrum* 67:3965–3972
- Tanaka S, Natsume M, Uchida M, Hotta N, Matsuda T, Spanut ZA, Hatsukade Y (2004) Measurement of metallic contaminants in food with a high-Tc SQUID. *Supercond Sci Technol* 17:620
- Tanaka S, Akai T, Hatsukade Y, Ohtani T, Suzuki S (2009) High Tc SQUID system for detection of small metallic contaminant in industrial products. *IEEE Trans Appl Supercond* 19: 882–885
- Tavarozzi I, Comani S, Del Gratta C, Di Luzio S, Romani GL, Gallina S, Zimarino M, Brisinda D, Fenici R, De Caterina R (2002) Magnetocardiography: current status and perspectives. Part II: clinical applications. *Ital Heart J* 3:151–165
- Thollon F, Burais N (1995) Geometrical optimization of sensors for eddy currents non-destructive testing and evaluation. *IEEE Trans Magn* 31:2026–2031
- Tse PW, Liu XC, Liu ZH, Wu B, He CF, Wang XJ (2011) An innovative design for using flexible printed coils for magnetostrictive-based longitudinal guided wave sensors in steel strand inspection. *Smart Mater Struct* 20:055001
- Tumanski S (2007) Induction coil sensors – a review. *Meas Sci Technol* 18:R31
- Tumanski S (2013) Modern magnetic field sensors – a review. *Organ* 89:1–12
- Urresty JC, Riba JR, Romeral L (2013) A back-emf based method to detect magnet failures in PMSMs. *IEEE Trans Magn* 49:591–598
- Wang P, Xiong L, Sun Y, Wang H, Tian G (2014) Features extraction of sensor array based PMFL technology for detection of rail cracks. *Measurement* 47:613–626
- Wang Z, Zhang S, Wan Z, Zhang M (2017) Analysis of non-destructive testing of three-dimensional braided composites based on SQUID. *Nondestr Test Eval* 32:21–35
- Wei G, Jianxin C (2002) A transducer made up of fluxgate sensors for testing wire rope defects. *IEEE Trans Instrum Meas* 51:120–124
- Weischedel HR, Ramsey RP (1989) Electromagnetic testing: a reliable method for the inspection of wire rope in the service. *NDT&E Int* 22:155–161
- Weisenstock H (1991) A review of SQUID magnetometry applied to nondestructive evaluation. *IEEE Trans Magn* 27:3231–3236
- Wikswo JP, Ma YP, Sepulveda NG, Tan S, Thomas IM, Lauder A (1993) Magnetic susceptibility imaging for nondestructive evaluation. *IEEE Trans Appl Supercond* 3:1995–2002
- Wu D, Liu Z, Wang X, Su L (2017) Composite magnetic flux leakage detection method for pipelines using alternating magnetic field excitation. *NDT&E Int* 91:148–155
- Xie M, Schneiderman JF, Chukharkin M, Kalabukhov A, Whitmarsh S, Lundqvist D, Winkler D (2015) High-Tc SQUID vs. low-Tc SQUID-based recordings on a head phantom: benchmarking for magnetoencephalography. *IEEE Trans Appl Supercond* 25:1–5
- Xu Z, Wu X, Li J, Kang Y (2012) Assessment of wall thinning in insulated ferromagnetic pipes using the time-to-peak of differential pulsed eddy-current testing signals. *NDT&E Int* 51:24–29

- Yang G, Zeng Z, Deng Y, Liu X, Udpa L, Tamburrino A, Udpa SS (2012) 3D EC-GMR sensor system for detection of subsurface defects at steel fastener sites. *NDT & E Int* 50:20–28
- Yang G, Dib G, Udpa L, Tamburrino A, Udpa SS (2015) Rotating field EC-GMR sensor for crack detection at fastener site in layered structures. *IEEE Sensors J* 15:463–470
- Ye C, Huang Y, Udpa L, Udpa S, Tamburrino A (2016) Magnetoresistive sensor with magnetic balance measurement for inspection of defects under magnetically permeable fasteners. *IEEE Sensors J* 16:2331–2338
- Ye C, Rosell A, Udpa L (2018a) Using Magnetoresistive sensors in nondestructive testing. *Mater Eval* 76:144–154
- Ye C, Wang Y, Tao Y (2018b) High-density large-scale TMR sensor array for magnetic field imaging. *IEEE Trans Instrum Meas.* <https://doi.org/10.1109/TIM.2018.2866299>
- Yu Y, Yan Y, Wang F, Tian GY, Zhang D (2014) An approach to reduce lift-off noise in pulsed eddy current nondestructive technology. *NDT&E Int* 63:1–6
- Yuan X, Li W, Chen G, Yin X, Ge J (2017) Circumferential current field testing system with TMR sensor array for non-contact detection and estimation of cracks on power plant piping. *Sensors Actuators A* 263:542–553
- Yuan X, Li W, Chen G, Yin X, Ge J, Yang W, Liu J, Ma W (2018) Inner circumferential current field testing system with TMR sensor arrays for inner-wall cracks inspection in aluminum tubes. *Measurement* 122:232–239
- Zeng Z, Liu X, Deng Y, Udpa L, Xuan L, Shih WCL, Fitzpatrick GL (2006) A parametric study of magneto-optic imaging using finite-element analysis applied to aircraft rivet site inspection. *IEEE Trans Magn* 42:3737–3744
- Zhang J, Yuan M, Song SJ, Kim HJ (2015) Precision measurement of coating thickness on ferromagnetic tube using pulsed eddy current technique. *Int J Precis Eng Manuf* 16:1723–1728
- Zhao Y, Li XM, Lin L, Lei MK (2011) Measurements of coating density using ultrasonic reflection coefficient phase spectrum. *Ultrasonics* 51:596–601
- Zhou D, Wang J, He Y, Chen D, Li K (2016) Influence of metallic shields on pulsed eddy current sensor for ferromagnetic materials defect detection. *Sensor Actuat A* 248:162–172
- Zhou D, Pan M, He Y, Du B (2017) Stress detection and measurement in ferromagnetic metals using pulse electromagnetic method with U-shaped sensor. *Measurement* 105:136–145



# Micromagnetic Materials Characterization

24

Klaus Szielasko and Ralf Tschuncky

## Contents

Introduction .....	882
Theoretical Background and Measured Effects .....	883
Magnetism and Magnetic Materials .....	883
Magnetic Hysteresis .....	883
Analysis of Upper Harmonics in the Tangential Magnetic Field .....	884
Magnetic Barkhausen Noise Analysis .....	884
Incremental Permeability .....	885
Eddy Current Impedance Analysis .....	886
Hybrid Approach .....	887
Calibration Procedures .....	887
Regression Analysis .....	888
Pattern Recognition .....	889
Validation Procedures .....	889
Validation Samples (Blind Samples) .....	890
Cross-validation Approaches .....	891
Understanding the Advantage of Multiparameter Approaches in Micromagnetic Materials Characterization .....	892
Multiparameter Advantage on the Example of High-Strength Steel .....	893
Multiparameter Advantage in Micromagnetic Microscopy .....	894
Summary .....	896
References .....	896

K. Szielasko (✉) · R. Tschuncky

Fraunhofer Institute for Nondestructive Testing IZFP, Saarbrücken, Germany  
e-mail: [klaus.szielasko@izfp.fraunhofer.de](mailto:klaus.szielasko@izfp.fraunhofer.de); [ralf.tschuncky@izfp.fraunhofer.de](mailto:ralf.tschuncky@izfp.fraunhofer.de)

**Abstract**

Micromagnetic materials characterization is receiving growing industrial acceptance and application due to significant improvements in sensor technology, data processing, and ease of use. The fundamental similarity between the interaction of microstructure with dislocations and magnetic domain walls is the basis of all micromagnetic approaches. This similarity leads to correlated interactions with magnetic and mechanical loads, resulting in, for example, the classical analogy between magnetic and mechanical hardness. In practical devices, a set of micromagnetic parameters is being determined in order to obtain a unique “fingerprint” of the material. In a calibration procedure, the multiparametric fingerprint is then mathematically related to target parameters such as hardness, hardening depth, strength, yield point, or residual stress. The multiparameter approach is preferred due to the fact that several material properties affect the magnetic behavior, so that a single measuring parameter will never be a unique function of a given target property. The main challenge is that sensor and part geometry are reflected in the magnetic parameter values, which makes it hard to collect calibration-relevant knowledge across several applications. Together with a growing variety of high-performance steel grades available today, this results in a need for individual, application-specific calibration. State-of-the-art micromagnetic testing systems address this issue by means of simplified, accelerated, and interactive calibration procedures and well-selected micromagnetic parameters of increased significance. The path pursued by developers today leads towards increasingly user-friendly devices with low calibration effort.

---

**Introduction**

Even though high-performance plastics, composites, and aluminum are gaining in importance, steel remains the most used material for machine parts and components today. Increasing demands of product quality on a constant, reliably high level in today’s automated production processes have become a strong driver for the development of advanced NDE solutions for the steel-producing and steel-processing industry. This concerns nondestructive sensing techniques for process control as well as quality assurance. Micromagnetic approaches play a major role in both domains, as far as the large group of ferromagnetic steels is concerned, which have magnetic domains interacting with microstructure and stress on a microscopic level. Ever since Heinrich Barkhausen discovered in 1919 that magnetic domain wall jumps can be detected inductively and envisioned that the noise produced by their seemingly random motions can be used to identify different steels (Barkhausen 1919), measurements of the magnetic hysteresis and domain phenomena have evolved into one of the most versatile sensing principles today.

## Theoretical Background and Measured Effects

### Magnetism and Magnetic Materials

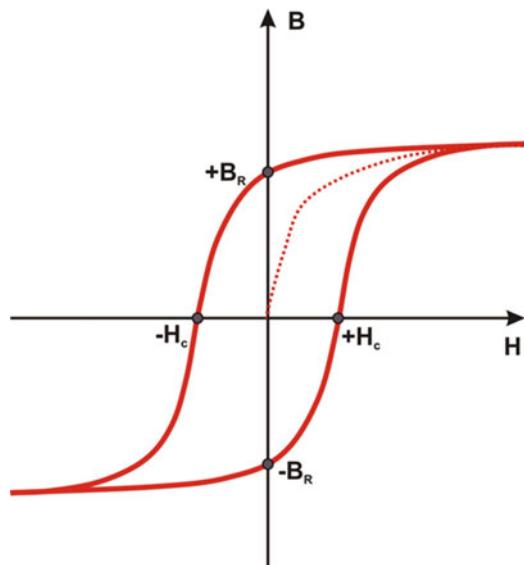
Electric and magnetic fields are fundamental physical phenomena connected to the spatial distribution of forces between charge-carrying matters. Maxwell's equations describe the origin and mutual interdependence of electric and magnetic fields (Maxwell 1865). Magnetic fields are created around moving charges (e.g., current flowing through a conductor) as well as by spin magnetic moments of elementary particles. Ferro- and ferrimagnetic materials contain areas of nonzero, aligned net magnetic moment, so-called magnetic domains (Weiss 1907). The formation of magnetic domains within ferro- and ferrimagnetic materials is an energetically preferred condition. The domain structure changes when the material is exposed to an external magnetic field: Domains with magnetization approximately in the direction of the external field grow, and others shrink. Magnetic domains of different orientation are separated by domain walls that energetically interact with micro-defects such as grain boundaries, dislocations, foreign atoms, and voids. Such defects impede the domain wall motion that is needed to change the size distribution of differently oriented domains, thus changing the overall magnetization of a material. Depending on density and size of micro-defects, materials can be hard or easy to be magnetized (and de-magnetized), which is why they are called hard-magnetic and soft-magnetic materials, correspondingly. The fact that microstructural defects affect both the mechanical and the magnetic material behaviors in similar ways is the basis of micromagnetic materials characterization.

### Magnetic Hysteresis

The magnetic domains are spontaneously magnetized to saturation, but the orientation of the magnetization vectors of the domains is statistically distributed across a ferromagnetic sample in the non-magnetized state in such a way that their overall effect neutralizes completely. Ferromagnetic materials tend to align the magnetization vectors of the domains toward the direction of any external magnetic field. In the case of an alternating external field magnetization of ferromagnetic materials, there is a complex relationship between the magnetic field strength  $H$  and the magnetic flux density  $B$ . In the case of periodic magnetization, a ferromagnetic material shows a hysteresis curve of  $B$  over  $H$  (see Fig. 1).

The hysteresis curve of a ferromagnetic material is traversed once every cycle during a periodically alternating field magnetization (after having passed the virgin curve until first saturation). During periodic magnetization, the flux density follows the hysteresis curve. The connection between mechanical and magnetic properties due to microstructural interactions of the Bloch walls is used for nondestructive electromagnetic testing. For example, the coercive force  $H_C$  often correlates well with the mechanical hardness.

**Fig. 1** Example of hysteresis curve (solid), virgin curve (dotted), coercive force  $H_c$ , and remanence flux density  $B_R$



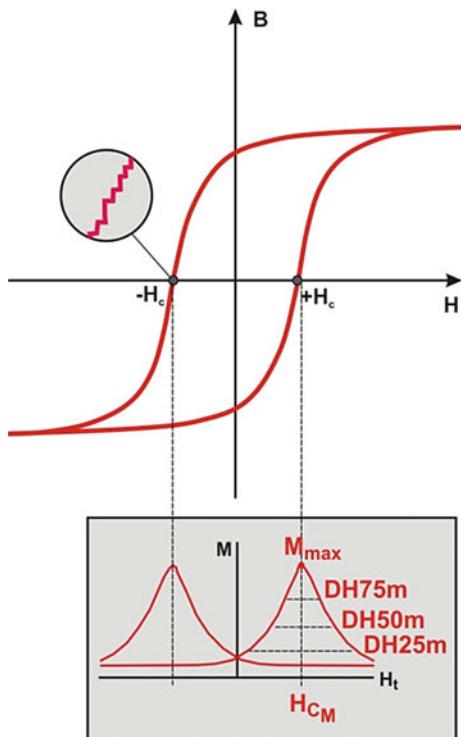
### Analysis of Upper Harmonics in the Tangential Magnetic Field

In the case of sinusoidal alternating field magnetization of a ferromagnetic material, the nonlinearity of the ferromagnetic hysteresis in the signal of the tangential field strength causes, in addition to the fundamental oscillation, pronounced upper harmonics caused by the nonlinear magnetic resistance of the measured object. Therefore, an analysis of the harmonic components occurring in the alternating magnetic field allows conclusions on the ferromagnetic properties of the material to be tested. The proportion of the upper harmonics in the tangential field can be determined by means of Fourier transformation. The extent of the upper harmonics depends on the design of the magnetization device applied and the material-specific behavior of the magnetic hysteresis of the material to be tested. This makes it possible to draw conclusions from the signals of the upper harmonics analysis on the microstructure of the material to be tested (Dobmann and Pitsch 1988; Pitsch 1990).

### Magnetic Barkhausen Noise Analysis

Magnetic Barkhausen noise analysis is based on a discovery by Heinrich Georg Barkhausen (Barkhausen 1919). The movement of the Bloch walls does not progress continuously but gradually during the magnetization process. This is due to the interaction of the Bloch walls with the microstructure of the ferromagnetic material. Bloch walls are pinned by lattice defects, and this pinning is gradually overcome under increasing external magnetic field (Kneller 1962; Cullity 1972). This results in sudden detachments of the Bloch walls and gradual increase of the domain size.

**Fig. 2** Schematic graph of the Barkhausen noise profile curve (bottom) with derived parameters in comparison with the magnetic hysteresis curve (top, section enlargement shows the gradual progress of the hysteresis caused by Barkhausen jumps)



The sudden detachment of a Bloch wall is referred to as Barkhausen jump. The discontinuous local magnetization changes are picked up inductively by a small coil as a noise signal of varying amplitude. The location of the greatest density of these pulses is, in terms of the magnetic hysteresis, in the vicinity of the coercive force  $H_C$ , since most of the Bloch wall jumps take place in this region. The envelope of the noise signal is determined by means of electronic and digital signal processing and finally recorded as a function of the magnetization field strength, resulting in the so-called Barkhausen noise profile curve  $M(H)$  (see Fig. 2). From this profile curve, different parameters for materials characterization can be derived (e.g., location of the maximum with respect to the field strength, etc.).

### Incremental Permeability

The method of incremental permeability analysis describes the reversible portion of the instantaneous magnetization change caused by small superimposed magnetic field excitation throughout the magnetic hysteresis loop. A practical measuring principle of the incremental permeability is based on eddy current impedance analysis. In the evaluation of the incremental permeability described here, a re-magnetization of the material is necessary. However, it is also possible to carry out evaluations of the incremental

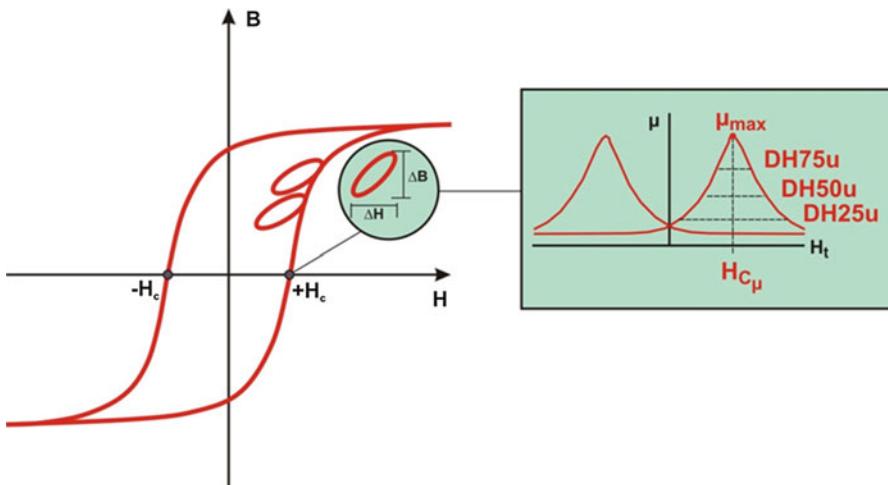
permeability in case of a constant field magnetization and superimposed alternating field magnetization. The operating point is determined by the constant field component within the magnetic hysteresis by passing through the superimposed alternating field an internal loop within the hysteresis. In this case, the value of the incremental permeability is defined by the ratio of the magnetic flux density change to the magnetic field strength change which arises as a result of the superimposed alternating field magnetization. This corresponds to the slope of the inner loop created by the superimposed alternating field. A qualitative measure for the incremental permeability over the entire hysteresis range is obtained by superimposing on the sinusoidal excitation field a higher-frequency magnetic alternating field such that an evaluation of the incremental permeability can take place over the entire course of the hysteresis.

This results in small internal loops of hysteresis – the hysteresis loops (see Fig. 3 on the left) whose surface converges to zero in purely reversible processes. The slope of these internal loops is a measure of the incremental permeability:  $\mu_\Delta = dB/dH$ . If the incremental permeability  $\mu_\Delta$  is plotted as a function of the field strength  $H$ , the result is the incremental permeability profile curve  $\mu(H)$  shown in Fig. 3 (right-hand side).

Different parameters for material characterization are derived from this profile curve (e.g., in analogy to the Barkhausen noise analysis, the location of the maxima with respect to the field strength, etc.).

## Eddy Current Impedance Analysis

The principle of the nondestructive eddy current impedance analysis is based on the fact that the impedance change of a coil is measured while interacting with the material to be tested (also see chapter on ► Chap 19, “Eddy Current Testing”).



**Fig. 3** Ferromagnetic hysteresis with hysteresis sub-loops (left) and the incremental permeability profile curve  $\mu(H)$  (right) derived from the slopes of the hysteresis loops

Circular, self-contained electric fields around the magnetic field lines occur due to the interaction of the material with the changing magnetic field caused by the eddy current testing coil. These fields cause the so-called eddy currents in the material (Heptner and Stroppe 1972). The intensity of the excited eddy currents depends on the frequency and amplitude of the alternating field as well as on the conductivity  $\sigma$  and permeability  $\mu$  of the sample. When a constant current amplitude is used, the voltage measured at the coil is proportional to the impedance and can be evaluated in the impedance plane. Microstructure affects the eddy currents through various interactions. Depending on the influences, different magnitude and phase changes of the coil impedance are caused (Heptner and Stroppe 1972).

## Hybrid Approach

In nondestructive evaluation, a hybrid approach is in general understood as a method which combines two or more nondestructive testing methods. Hybrid methods or approaches acquire several measuring quantities with different physical information content of the tested materials. In addition, a hybrid approach has different sensitivity against disturbing influences, and, moreover, the option for different analysis depths exists. In many cases, the measured quantities of hybrid methods allow for a quantitative determination of material properties like hardness, hardening depth, residual stresses, etc. using calibration procedures. Another advantage of hybrid approaches is that different material features can be predicted in some applications simultaneously (Tschuncky et al. 2016).

In the area of nondestructive evaluation, a lot of possible hybrid combinations exist. Because of this the focus of this paragraph is in the area of electromagnetic hybrid approaches only. Therefore, only ferromagnetic materials can be tested with such electromagnetic hybrid approaches.

The 3MA approach is a rather well-known electromagnetic hybrid method. 3MA is an acronym for Micromagnetic, Multiparameter, Microstructure, and Stress Analysis (Altpeter et al. 2002). This approach combines different electromagnetic methods and acquires several electromagnetic parameters. Based on these parameters and using a calibration procedure, the quantitative determination of material properties for ferromagnetic materials becomes possible.

---

## Calibration Procedures

In order to be able to quantitatively determine mechanical technological characteristics such as hardness, tensile strength, or residual stresses, the so-called target values, it is necessary to calibrate electromagnetic nondestructive testing systems. This applies in particular for hybrid systems in order to be able to use their large number of measuring quantities in a targeted manner.

During calibration procedures, the corresponding target values are assigned to the electromagnetic measuring quantities. This requires sample sets with conventionally

determined reference data, which can be used to acquire electromagnetic measurement quantities on samples with known target values during calibration procedures.

Several methods are available as mathematical tools for determining correlations between the measuring quantities and the material characteristics. For example, the regression analysis, which analyzes the calibration data set by means of regression calculation with regard to the significance of the individual electromagnetic measurement quantity, takes into account the measurement quantities values in different powers, as reciprocal values and in the form of selected products and quotients. Another option is to use pattern recognition methods/algorithms, which analyze the measurement quantities values of the sample to be characterized based on similarity considerations to the measurement quantities values of the calibration samples and thereby calculate the assigned target values (Stork 2001).

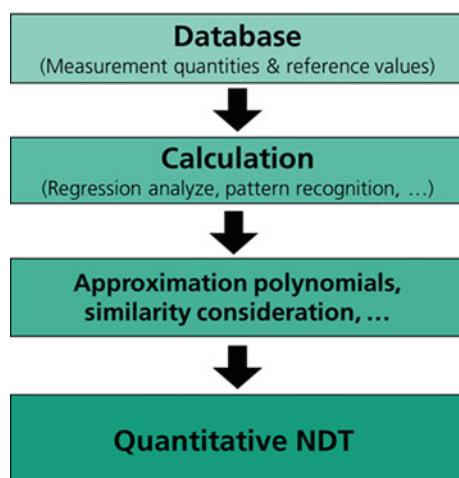
Figure 4 shows schematically the procedure of calibration. It becomes clear that, regardless of the calibration method, the measurement quantities space taken up at the calibration sample set always represents the basis for the calibration procedures.

## Regression Analysis

Regression analysis is a statistical analysis method that aims to establish relations between one dependent and one or more independent variables. In this case, the relation between the electromagnetic measurement quantities and the target value(s) is made available to this mathematical aid. The data matrix consisting of the electromagnetic measurement quantities is interpreted as a plurality of independent variables and the target variable vector corresponding to the measurement quantities as the dependent variable, so that the regression analysis should find correlations between the measurement quantities and the associated result vectors.

The regression analysis models the correlation by a polynomial that captures the measurement quantities – target value(s) relation(s). The approximate optimal

**Fig. 4** Schematic diagram of calibration procedures



coefficients of the regression polynomial are determined using regression analysis on the basis of data from calibration measurements, which are analyzed for the significance of the individual electromagnetic measurement quantities. The calibration measurements are taken on a graded calibration sample set, which must cover the target value range under consideration.

When applying linear regression analysis for the calibration of electromagnetic testing methods, a polynomial is parameterized, which in subsequent measurements from the measurement quantities values approximates the expected target value.

## Pattern Recognition

In general, pattern recognition is always model-free, since here only the similarity of the measurement quantities values with the measurement quantities values contained in the calibration database is determined to implement the calibration. It identifies those samples from the calibration database from which the measured sample deviates the least.

A method for this may be a nearest neighbor search, which performs a consideration of the Euclidean distances of the measurement quantities values to determine the similarity (Tschuncky 2011). Nearest neighbor search is a nonparametric method. This method does not make any assumptions about the parametric form of the underlying distribution of the test quantities or does not start from a model to determine the target size test set context.

In general, a feature extraction is first performed on a nearest neighbor search. Here, from the set of available measurement quantities, those which are to be used for pattern recognition are selected (Stork 2001).

After the measurement quantities to be used have been determined and normalized, an evaluation of the measurement quantities values of the sample to be characterized with respect to the Euclidean distance to the corresponding entries in the calibration database takes place. There is a difference in the formation for each individual normalized measurement quantities of the calibration database and the corresponding normalized measurement quantities of the sample to be characterized. These differences are added up to a normalized total distance assigned to the corresponding calibration data set. After the total distance is formed, the samples from the calibration database are referred to as nearest neighbors of the unknown sample if their Euclidean distance is below a user-selected bound (maximum Euclidean distance). These nearest neighbors are used to generate/calculate the result of pattern recognition (target value of the unknown sample) (Tschuncky 2011).

---

## Validation Procedures

The goal of validation procedures is to quantify and test the accuracy, reliability, and repeatability to be expected from the calibrated nondestructive system under application conditions. Producing reference samples of well-defined conditions is a significant challenge, because involving realistic process deviations is crucial if the

validation results should be representative of the final application. This circumstance raises both effort and costs of reference sample production significantly and explains the need for strategic approaches of validation in order to keep the need for well-controlled samples within reasonable limits.

## Validation Samples (Blind Samples)

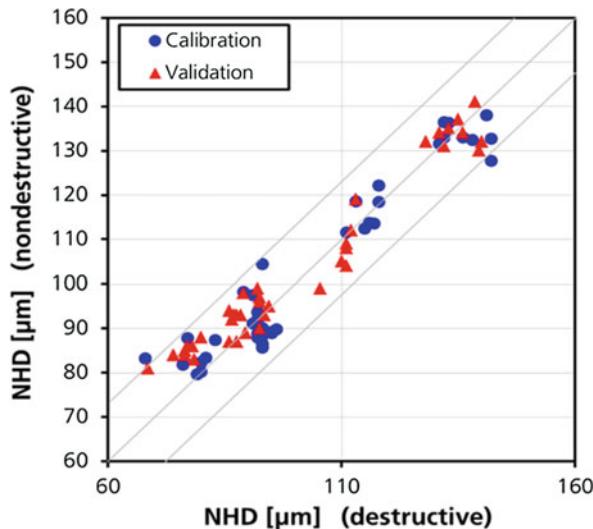
A straightforward validation approach is to evaluate the calibration on a sufficiently large set of samples that were not part of the calibration procedure. On the one hand, these samples should cover the same range of the target parameters as the validation samples did. On the other hand, these samples must never be identical or very close to the calibration samples, because higher order polynomials and pattern recognition approaches tend to classify calibration data more precisely than actual validation data.

Figure 5 shows an example of calibration and validation with large sets of samples in both scenarios. In this example, the nitride hardening depth (NHD) at the tribologically relevant surface of piston rings was evaluated by means of magnetic Barkhausen noise (MBN) analysis and a miniature 3MA device (Szielasko 2009; Szielasko et al. 2013). Using regression analysis on a set of calibration data from 40 samples, a polynomial function was determined that approximated the NHD by a mathematical function of MBN parameters. This function was then applied to data measured on another 40 samples. Since both sample sets were rather large, the root mean standard error (RMSE) of calibration and validation data were quite similar (around 7.3 µm). As long as the samples used can be considered representative for actual production, achieving similar RMSE values in calibration and validation indicates that these are representative for real-life application. The measuring error of the system can therefore be expected to lie within 95% – confidence bounds of  $\pm 2 \cdot \text{RMSE} = 14.6 \mu\text{m}$  (indicated by the thin solid lines in Fig. 5).

If the RMSE of the validation is significantly higher than the RMSE of the calibration, one out of the following reasons might be responsible:

- Insufficient amount of calibration samples – The calibration sample set is not representative for application conditions; therefore several validation samples have magnetic properties deviating from calibration knowledge.
- Differently biased distribution of reference values in calibration and validation data – This may occur if validation and calibration cover different ranges of the target parameter, leading to a need for extrapolation of the calibration (which typically fails in case of higher-order polynomials or in case of pattern recognition). In order to avoid this scenario, calibration and validation samples have to be selected in such a way that they cover the same target range without bias.
- Material differences between calibration and validation samples – This occurs in case of lot variations or process parameter scattering between calibration and validation. This scenario can be avoided by collecting sample sets over an extended period of time, so that these influences are present within both calibration and validation data.

**Fig. 5** Calibration and validation data of nitride hardening depth (NHD), determined using magnetic Barkhausen noise parameters and from micrographs (Szielasko 2009)



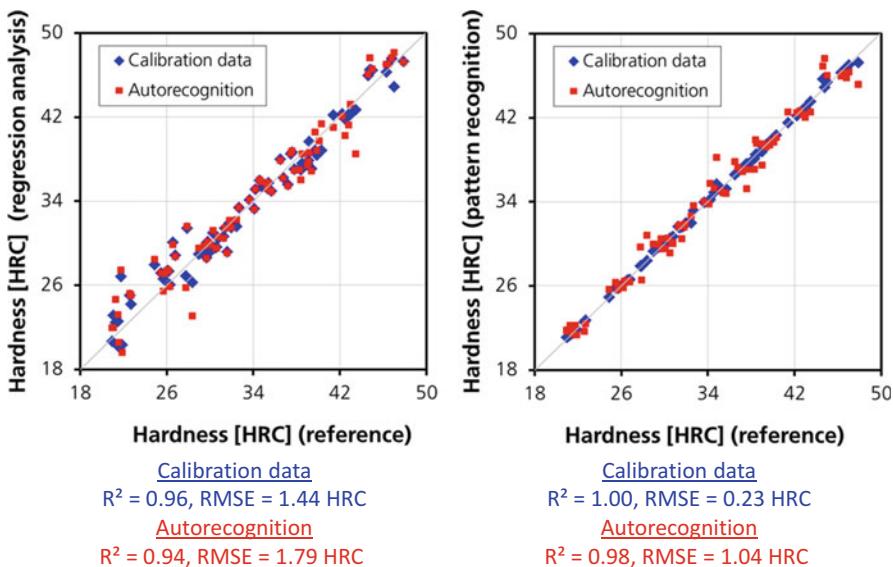
In either case, a comparison of both RMSE values is a suitable measure for a final acceptance test of the calibration.

### Cross-validation Approaches

Under real-life conditions, high costs for production and characterization of well-defined reference material lead to a rather low amount (typically in the order of 10 or 20) of reference samples. On the one hand, splitting up this sample set into two equally sized groups drastically increases the risk of incomplete calibration, underrepresenting the range of target values. On the other hand, if the validation set is much smaller than the calibration set, the outcome of the validation might not be statistically representative for later application. Validating the calibration with calibration data always results in the system performance appearing better than in later practical conditions. This is particularly important if pattern recognition is used.

This issue is commonly addressed by means of cross-validation approaches. In Szielasko (2009), a so-called autorecognition test is discussed. It maximizes both calibration and validation data sets by varying the selection of just one validation sample at a time, using all the others for calibration. The target values predicted in this validation procedure are highly representative for later application, because in each calibration scenario, almost all samples were used for calibration, and at the same time, each of the samples was used as validation sample in different calibrations. This situation is highly similar to using all samples for calibration and inspecting a totally unknown sample – therefore, this procedure is recommended for assessing the real-life performance of the system.

A comparison of validation with calibration data and validation via auto-recognition test is given in Fig. 6, for the scenarios of pattern recognition and



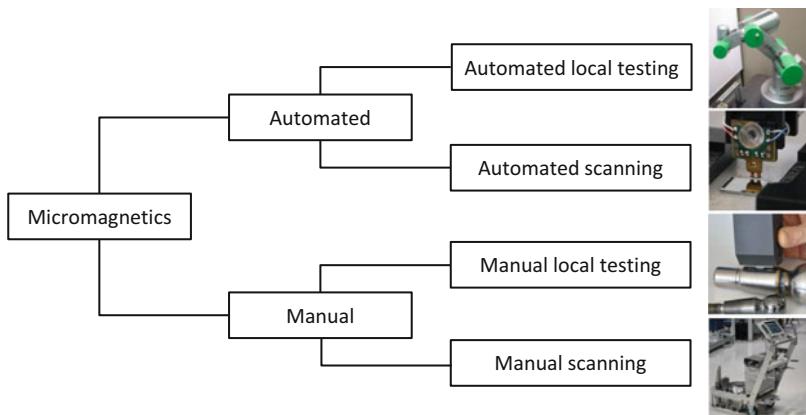
**Fig. 6** Comparison of validation with calibration data and validation using the leave-one-out “autorecognition” test (for pattern recognition to the left and for regression analysis to the right)

regression analysis (Szielasko 2009). Comparing between pattern recognition and regression analysis indicates that pattern recognition should deliver superior performance in this application. Without the autorecognition test, pattern recognition performance appears unrealistically high ( $R^2 = 1.00$ , RMSE = 0.23 HRC), because calibration data always contains perfect matches in the nearest neighbor search. In the autorecognition test, each test data point was excluded from the calibration at a time, leading to a performance close to reality.

### Understanding the Advantage of Multiparameter Approaches in Micromagnetic Materials Characterization

Micromagnetic materials characterization faces different challenges, depending on the field of application. Figure 7 provides an overview of different automated and manual modes of operation.

Automated testing is generally associated with higher importance of wear protection, lift-off control, and speed optimization. In manual local testing, frequent challenges lie in designing a part-specific sensor, reaching an acceptable lifetime, and being able to reproduce the sensor characteristics in case of repair or replacement. If surface areas have to be scanned, the use of sensor arrays and fast multichannel devices can be considered. Sensor and channel matching can be challenging and may require accurate component selection for electronics and sensor materials.



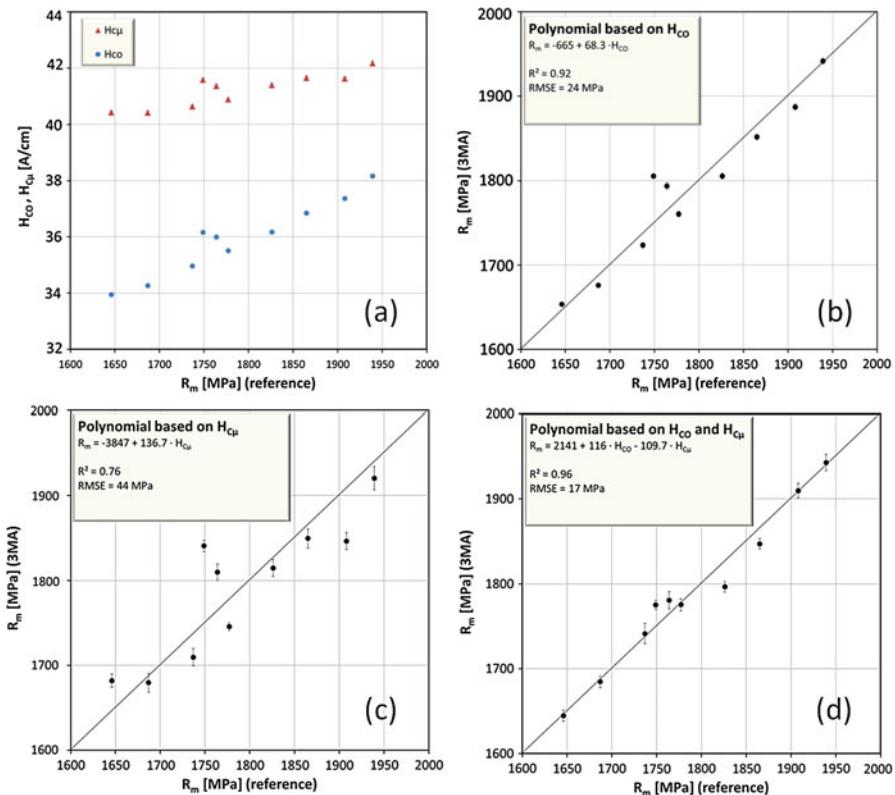
**Fig. 7** Micromagnetic materials characterization, divided into automated and manual operation modes

In all scenarios mentioned above, multiparametric approaches excel through their advantage of compensating superimposed disturbances by combining the information contained in different micromagnetic parameters. The following sections illustrate two examples, one regarding macroscopic properties and the other regarding microscopy.

### Multiparameter Advantage on the Example of High-Strength Steel

In the following application, the ultimate strength  $R_m$  of 1.4310 steel in different cold-worked conditions and across a thickness range of 0.4–0.5 mm was determined nondestructively using 3MA technology (Szielasko et al. 2014). The tensile strength ranged from 1650 to 1950 MPa and was determined with a small air gap between sensor and sample.

In order to demonstrate the advantage of multiparameter analysis, we focus on the determination of two coercivity values, one determined in harmonics analysis of magnetic tangential field strength ( $H_{CO}$ ) and the other one determined in eddy current based incremental permeability ( $H_{C\mu}$ ). Due to different physical interactions, both coercivity values differ slightly, as shown in Fig. 8a. Calibrating the device with regression analysis using just one out of the two coercivity parameters leads to significant outliers between 1750 and 1800 MPa (Fig. 8b, c). The reason is that each of the coercivities shows a deviating trend as a function of strength in this interval. Combining both coercivities in one polynomial clearly decreases the effect of these outliers, because the parameters reflect the outliers differently, thus leading to a suppression of their effect on the result (Fig. 8c). In most practical applications, up to 10 micromagnetic parameters are used in order to reach sufficient accuracy and robustness. An obvious risk is that using too high a number of parameters for calibration can lead to an overfit that does not represent the general trend but rather



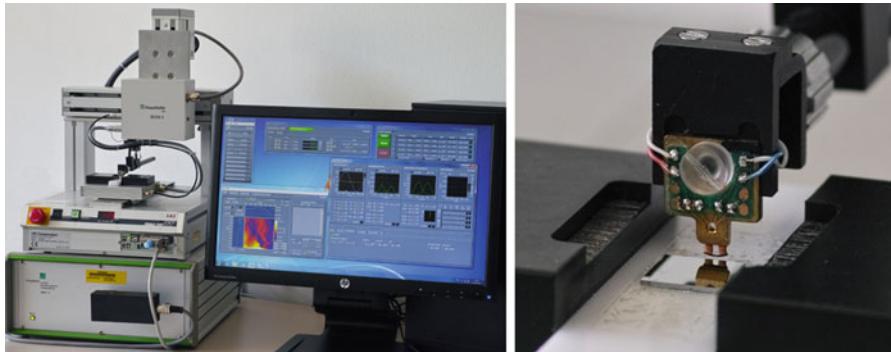
**Fig. 8** Advantage of multiparameter analysis: Based on the parameters  $H_{CO}$  and  $H_{C\mu}$  (a), polynomials are determined that reflect ultimate strength  $R_m$ . Graph (c) shows the results using  $H_{CO}$  only and (c) using  $H_{C\mu}$  only. Combining both parameters in one polynomial significantly increases the accuracy (d)

exactly the training data. This undesired scenario is avoided by applying statistical tests to the data and restricting the number of parameters according to the amount of calibration data available.

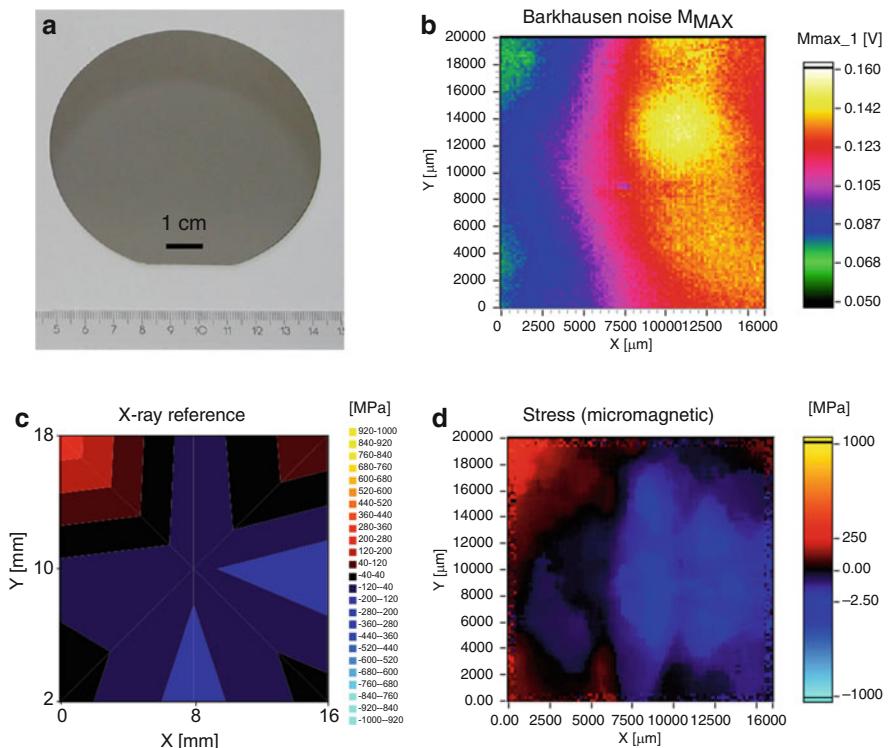
### Multiparameter Advantage in Micromagnetic Microscopy

The deposition of thin films on wafers, e.g., for magnetic sensor manufacturing, is often associated with significant local stress fields that may, in the worst case, lead to delamination of the film. In the following application,  $\text{SiO}_2$  wafers were coated with NiFe-Cu-NiFe stacks (200 nm–50 nm–200 nm) for manufacturing giant magnetoimpedance sensors. The stress homogeneity across the wafer affects the sensitivity of the sensors separated from them later on. As NiFe is ferromagnetic, the

dependence of the micromagnetic parameters on stress was used to determine the stress quantitatively. A Barkhausen Noise and Eddy Current Microscope (BEMI, Fig. 9) was used for local micromagnetic multiparameter analysis at a lateral resolution of about 10–20  $\mu\text{m}$  (Szielasko et al. 2004). The operation principle of



**Fig. 9** Barkhausen noise and eddy current microscope, BEMI



**Fig. 10** Wafers investigated with BEMI (a), magnetic Barkhausen noise amplitude  $M_{MAX}$  (b), reference measurement with X-ray diffraction (c), and calibrated result with five terms (d)

BEMI is similar to scanning Hall probe microscopy (Chang et al. 1992). However, using the 3MA technique and a miniaturized magnetoinductive probe, BEMI obtains significantly more materials information compared to Hall probe microscopy. A few points across the surface were investigated with X-ray diffraction for reference and calibration.

Figure 10a shows one of the wafers investigated, and Fig. 10b represents the inhomogeneous distribution of the maximum magnetic Barkhausen noise amplitude ( $M_{MAX}$ ) across the center area of the wafer. In Fig. 10c, the results of X-ray diffraction analysis are plotted. The device was calibrated using regression analysis with five terms, including  $M_{MAX}$ . Figure 10d illustrates that the resulting polynomial, plotted for each pixel of the scan, represents a much finer view of local stress than X-ray does and shows more information than  $M_{MAX}$  alone as well. Here, the multiparameter approach allows for a better fit between micromagnetically determined stress and reference values as well as a better view of local details in the stress distribution.

---

## Summary

Micromagnetic materials characterization is a versatile tool for materials development, semifinished goods inspection, and part quality assurance in the steel-producing and steel-processing industry. Multiparameter micromagnetic approaches have the advantage of better disturbance suppression as well as generally wider field of application as compared to single-parameter approaches. All micromagnetic approaches have to be calibrated, and there are mathematical solutions readily implemented in order to validate the accuracy to be expected in the later application. Micromagnetic analysis systems and sensors are usually optimized for specific applications, which range from microscopy up to high-speed strip steel inspection.

---

## References

- Altpeter I, Becker R, Dobmann G, Kern R, Theiner WA, Yashan A (2002) Robust solutions of inverse problems in electromagnetic non-destructive evaluation. *Inverse Problems* 18:1907–1921
- Barkhausen H (1919) Zwei mit Hilfe der neuen Verstärker entdeckte Erscheinungen. *Phys Z* 20:401–403
- Chang AM, Hallen HD, Harriott L, Hess HF, Kao HL, Kwo J, Miller RE, Wolfe R, Van Der Ziel J, Chang TY (1992) Scanning hall probe microscopy. *Appl Phys Lett* 61(16):1974. <https://doi.org/10.1063/1.108334>
- Cullity BD (1972) Introduction to magnetic materials. Addison-Wesley, Reading
- Dobmann G, Pitsch H (1988) Verfahren zum zerstörungsfreien Messen magnetischer Eigenschaften eines Prüfkörpers sowieVorrichtung zum zerstörungs-freien Messen magnetischer Eigenschaften eines Prüfkörpers. German patent DE3037932A1, April 23, 1988
- Heptner H, Stroppe H (1972) Magnetische und magnetinduktive Werkstoffprüfung. VEB Deutscher Verlag für Grundstoffindustrie, Leipzig
- Kneller E (1962) Ferromagnetismus. Springer, Berlin

- Maxwell JC (1865) A dynamical theory of the electromagnetic field. *Philos Trans R Soc Lond* 155:459–512
- Pitsch H (1990) Die Entwicklung und Erprobung der Oberwellenanalyse im Zeitsignal der magnetischen Tangentialfeldstärke als neues Modul des 3MA-Ansatzes. Doctoral dissertation, Saarland University, Sarbrücken
- Stork D (2001) Pattern classification. Wiley, New York
- Szielasko K (2009) Development of metrological modules for electromagnetic multiparameter materials characterization and testing. Doctoral dissertation, Saarland University, Saarbrücken (in German)
- Szielasko K, Kopp M, Tschuncky R, Lugin S, Altpeter I (2004) Barkhausenrausch- und Wirbelstrom mikroskopie zur ortsaufgelösten Charakterisierung von dünnen Schichten. DGZfP annual conference 2004 V13
- Szielasko K, Mironenko I, Altpeter I, Herrmann HG, Boller C (2013) Minimalistic devices and sensors for micromagnetic materials characterization. *IEEE Trans Magn* 49(1):101–104
- Szielasko K, Kopp M, Tschuncky R and Herrmann HG (2014) Zerstörungsfreie Bestimmung von Werkstoffeigenschaften mit mikromagnetischen Multiparameter-Prüfverfahren. In: *Werkstoffe in der Fertigung* 1/2014, pp 45–46. ISSN 0939-2629/B 25800
- Tschuncky R (2011) Sensor- und geräteunabhängige Kalibrierung elektromagnetischer zerstörungsfreier Prüfverfahren zur praxisorientierten Werkstoffcharakterisierung. Doctoral dissertation, Saarland University, Saarbrücken
- Tschuncky R, Szielasko K, Altpeter I (2016) Hybrid methods for materials characterization. In: Hübschen G, Altpeter I, Tschuncky R, Herrmann HG (eds) *Materials characterization using nondestructive evaluation (NDE) methods*. Woodhead Publishing, Cambridge, pp 263–291
- Weiss P (1907) L'Hypothèse du Champ Moléculaire et de la Propriété Ferromagnétique. *J Phys* 6:661–690



# Electrochemical Techniques, Impedance, and Spectroscopy

25

Shengxi Li, Qixin Zhou, and Hongbo Cong

## Contents

Basics of the Method .....	900
Instrumentation .....	901
Typical Applications .....	902
Atmospheric Corrosion .....	902
Corrosion in Concrete .....	905
Coatings and Films .....	908
Inhibitors .....	913
High-Temperature Applications .....	915
Other Applications .....	916
Summary .....	917
References .....	917

## Abstract

Electrochemical impedance spectroscopy (EIS) is a powerful technique for characterizing a wide range of electrochemical systems. It is based on the application of a small amplitude, alternating current signal to the systems, and therefore is considered a non-destructive technique. This chapter begins with a brief introduction of the fundamentals of EIS, then focuses on instrumentation characteristics, especially the various electrode configurations, and finally discusses the wide applications of EIS in corrosion studies. The applications of EIS in the following corrosion topics are reviewed: atmospheric corrosion, corrosion in concrete, coatings and films, inhibitors, high-temperature corrosion, oil and gas industry-related corrosion, microbiologically influenced corrosion, and stress corrosion

---

S. Li · Q. Zhou · H. Cong (✉)

Department of Chemical and Biomolecular Engineering, Corrosion Engineering Program,

The University of Akron, Akron, OH, USA

e-mail: [sli17@uakron.edu](mailto:sli17@uakron.edu); [qzhou@uakron.edu](mailto:qzhou@uakron.edu); [hcong@uakron.edu](mailto:hcong@uakron.edu)

cracking. For each application, besides the review of the capability of EIS, typical equivalent circuit models are also highlighted. In some applications, the limitations of EIS measurements are also pointed out.

## Basics of the Method

Electrochemical impedance spectroscopy (EIS) is one of the most powerful and versatile techniques for corrosion studies. It can be used for corrosion rate measurements, corrosivity monitoring, coating integrity measurements, and reaction mechanism investigations. EIS is usually performed by applying a continuous alternating current (AC) signal to a steady-state electrochemical system and then measuring the current response. Because the AC signal perturbation is a small excitation signal, EIS is essentially a nondestructive technique.

When a small sinusoidal voltage (Eq. 1) is applied to a system, a sinusoidal current (Eq. 2) is observed. The resulting phase shift is shown in Eq. 3.

$$U(t) = U_m \sin(\omega t + \varphi_u) \quad (1)$$

$$I(t) = I_m \sin(\omega t + \varphi_i) \quad (2)$$

$$\varphi = \varphi_u - \varphi_i \quad (3)$$

The magnitude of the amplitudes of perturbation and response,  $U_m$  and  $I_m$ , and the phase shift  $\varphi$  depend on the reactions taking place in the systems (Macdonald 1987). The perturbation and response are transformed from a function of time into the frequency domain using the Laplace transformation. The ratio of the Laplace-transformed voltage and current  $\Delta U(s)/\Delta I(s)$  defines the frequency-dependent electrochemical impedance  $Z(s)$  (Rammelt and Reinhard 1992).

$$Z(s) = Z(j\omega) \quad (4)$$

where  $s$  is the Laplace variable,  $\omega = 2\pi f$  and  $j = \sqrt{-1}$ .

The impedance  $Z(j\omega)$  is a complex number which can be represented in Cartesian coordinate (Eq. 5).

$$Z(j\omega) = Z' + jZ'' \quad (5)$$

$|Z|$  is the modulus of the impedance, and  $Z'$  and  $Z''$  are the real and imaginary parts of the impedance, respectively. The relationships between these quantities are:

$$|Z| = \sqrt{(Z')^2 + (Z'')^2} \quad (6)$$

$$Z' = |Z| \cos \varphi \quad (7)$$

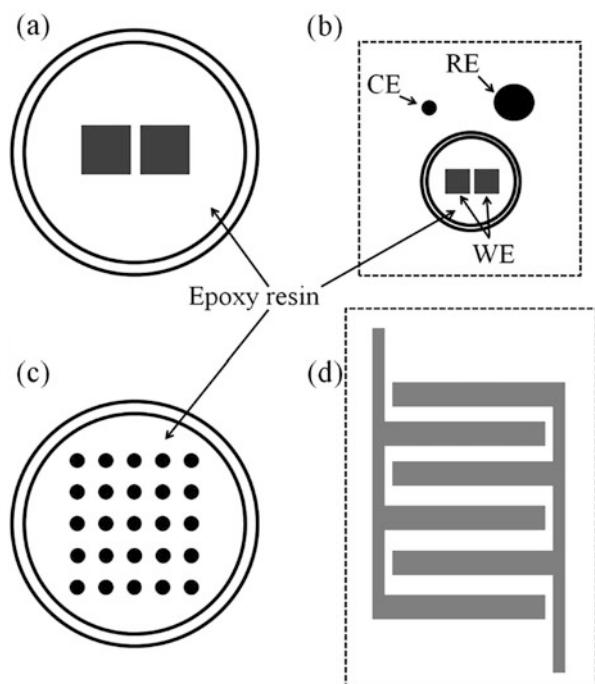
$$Z'' = |Z| \sin \varphi \quad (8)$$

The plots of  $Z(j\omega)$  measured at different frequencies are called Nyquist plots, complex plots, impedance diagrams, or impedance spectra. The plots showing the logarithm of the magnitude of the impedance ( $\log |Z|$ ) and the phase shift ( $\varphi$ ) as a function of the logarithm of the frequency ( $f$ ) are called Bode plots.

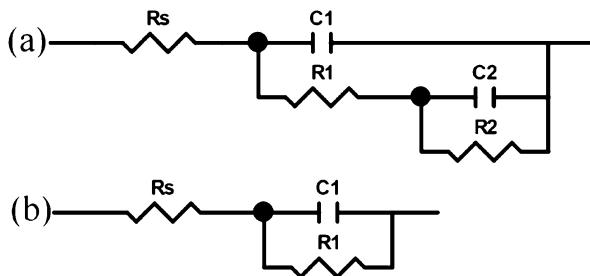
## Instrumentation

The conventional EIS measurements are usually conducted in three electrode systems, with the test materials as the working electrodes. The counter electrodes which have to be inert and do not participate in the electrochemical reaction can be made of Pt, Au, or graphite. The common reference electrodes include saturated calomel electrode (SCE) and Ag/AgCl electrode. However, in many applications such as under thin electrolyte films or in high-temperature environments, the conventional reference electrodes do not function properly. Instead, electrode systems without the conventional reference electrodes should be used. An example is the two-electrode system (Fig. 1a) which usually consists of two identical electrodes made of the test materials and has been widely used in atmospheric corrosion monitoring (Nishikata et al. 1995a). Figure 1b shows a modification of this two-electrode system that was used for high-temperature corrosion monitoring (Aung and Liu 2012). Interdigitated electrodes (comb-like electrodes) are another two-electrode system which have been

**Fig. 1** Schematic diagrams showing examples of electrode systems for corrosion monitoring using EIS (Nishikata et al. 1995a; Aung and Liu 2012; Cooper et al. 2006; and El-Mahdy and Kim 2004)



**Fig. 2** (a) Beaunier's equivalent circuit model (Beaunier et al. 1976); (b) simplified Randles equivalent circuit model (Randles 1947)



used to effectively monitor atmospheric corrosion (Fig. 1d) (El-Mahdy and Kim 2004). Finally, multielectrode array (MEA) was also successfully used for EIS monitoring (Fig. 1c) (Cooper et al. 2006).

The analysis of impedance data is often based on the classical equivalent circuit (Fig. 2a) proposed by Beaunier et al. (1976) and its variants, for example, the simplified Randles equivalent circuit shown in Fig. 2b. Beaunier's circuit consists of an uncompensated electrolyte resistance ( $R_s$ ) in series with a parallel connection of a capacitance ( $C_1$ ) and a branch including the pore solution resistance ( $R_1$ ) in series with a parallel connection of a double-layer capacitance ( $C_2$ ) at the coating/substrate interface and a resistance ( $R_2$ ) representing the charge transfer process (i.e., corrosion). In the variants of Beaunier's circuit, additional elements such as constant phase element (CPE), inductance ( $L$ ), and Warburg resistance ( $W$ ) are usually used to replace either resistance or capacitance. Substituting CPE for capacitance may greatly improve the quality of the fitting between the experimental data, but tends to create ambiguities in the physical interpretation of the results. This is because the CPE modulus cannot be simply identified with the capacitance and calculation of the effective capacitance from the CPE parameters requires a detailed knowledge on the physical reasons for the CPE behavior (Hirschorn et al. 2010).

Besides the commonly used and relatively simple equivalent circuit models, complicated physical models were sometimes used to interpret EIS data obtained from sophisticated systems. An example is the transmission line (TML) model, which was first developed by de Levie in his work on porous electrodes (de Levie 1963). The TML model and its variations have been used in many cases for EIS data analyses, such as EIS obtained from atmospheric corrosion under thin electrolyte films (Nishikata et al. 1995a) and from samples subjected to stress corrosion cracking (Oskui et al. 2012).

## Typical Applications

### Atmospheric Corrosion

Atmospheric corrosion is an electrochemical process which occurs typically under thin electrolyte layers, with or without salts and gaseous pollutants dissolved in them. It was found that the atmospheric corrosion rate of metals depends on the

thickness of the thin electrolyte layer. A change in the electrolyte film thickness affects the oxygen transport rate through the electrolyte layer and the dissolution of corrosion products, which influence the cathodic reaction rates (in neutral and alkaline solutions) and anodic process, respectively.

The thin electrolyte films present challenges in the corrosion monitoring using conventional electrochemical techniques, because the solution resistance becomes exceptionally high when electrolyte films are very thin. This situation results in an extremely high ohmic potential drop and a non-uniform current distribution, which renders the corrosion rate measurement problematic (Nishikata et al. 1995a). However, EIS can be used to measure the corrosion rates of metals because the solution resistance is considered during the measurements. The solution resistance is estimated from the impedance measured in high-frequency range of the EIS spectra, while the impedance in the low-frequency range provides the sum of polarization resistance ( $R_p$ ) and solution resistance ( $R_s$ ). The calculated polarization resistance is then converted to the corrosion rates of the metals.

Tsuru and Nishikata et al. reported extensive studies on EIS monitoring of the corrosion rates of metals under thin electrolyte films (~10–1000  $\mu\text{m}$ ) either during laboratory-simulated wet-dry cycles or under outdoor exposure conditions (Nishikata et al. 1995a, b, 1997, 2005; Vera Cruz et al. 1996; El-Mahdy et al. 2000a, b; Yadav et al. 2004a, b). In the majority of their work, they used two or three electrodes embedded in parallel in epoxy resin (Fig. 1a) to monitor the thin electrolyte film-induced corrosion rates of metals, including stainless steels, zinc, galvanized steels, and copper. For both laboratory and outdoor exposure (Nishikata et al. 1995b; EL-Mahdy and Kim 2004) experiments, the electrolyte was pre-deposited onto the electrode surface as a drop or film. The impedance measurements were conducted by either acquiring the EIS spectra in the frequency range of 10 mHz to 10 kHz (Nishikata et al. 1995a, 1997) or continuously monitoring the impedance at fixed frequencies (i.e., 10 mHz and 10 kHz) (Nishikata et al. 1995b; Vera Cruz et al. 1996; El-Mahdy et al. 2000a, b; Yadav et al. 2004a, b). For all these measurements, the excitation voltage had an amplitude of 10 mV. By investigating the full EIS spectra, Nishikata et al. found that the data could be fitted using a 1-D distributed constants type equivalent circuit called the TML model (Nishikata et al. 1995a). The TML analysis suggested that the EIS measurements provided information on the current distribution on metals covered by thin electrolyte films. When the phase angle exceeded  $-45^\circ$ , the current distribution was uniform and accurate corrosion rates were obtained. However, when phase angle values went no further than  $-45^\circ$ , an uneven current distribution needed to be considered. Using the TML analysis, the same group also successfully estimated charge transfer resistance of steel under uneven current distribution (Shi et al. 2015).

In addition to the model analysis of EIS, Tsuru and Nishikata et al. focused their research on monitoring corrosion rate under thin electrolyte films through fixed-frequency (i.e., 10 mHz and 10 kHz) impedance measurements. The solution resistance ( $R_s$ ) was obtained from the impedance measured at 10 kHz ( $Z_H$ ) while the polarization resistance ( $R_p$ ) was obtained by subtracting the impedance at 10 mHz ( $Z_L$ ) from  $Z_H$ . Corrosion rate was then calculated from the polarization

resistance. Different corrosion behavior was observed under the thin electrolyte layers for different metals. For example, the corrosion rate of copper was independent of the electrolyte film thickness when it was covered with a pH 6 Na<sub>2</sub>SO<sub>4</sub> solution, but reached a maximum value at a few tens of micrometers when the electrolyte was acidic (i.e., pH 3 and 4) (Nishikata et al. 1995a). The corrosion rate of iron under electrolyte layers of different thickness (i.e., 10–1000 μm) and pH levels (i.e., 3.0, 4.0, and 5.7) showed a maximum when the electrolyte thickness was in the range of 20–30 μm in all three pH levels (Nishikata et al. 1997). For a 55%Al-Zn alloy coated steel, it was found that the corrosion rate decreased substantially as the electrolyte thickness increased from 15 to 100 μm and did not change significantly with a further increase to 1000 μm (El-Mahdy et al. 2000a). They also monitored pitting corrosion of stainless steel under chloride-containing electrolyte films (Vera Cruz et al. 1996). It was found that under wet-dry cyclic condition, predominant pit generation and growth occurred immediately before complete drying of the metal surface.

Tsuru and Nishikata et al. also employed pairs of comb-shaped (i.e., interdigitated) electrodes (Fig. 1d) for their atmospheric corrosion studies using EIS (Nishikata et al. 2005, 2014). They monitored the impedance data at both 10 kHz and 10 mHz. The corrosion rates were obtained from the difference between the two impedance values (Nishikata et al. 2005). It was discovered that the reciprocal of average impedance at the low frequency (10 mHz) correlated well with the average corrosion rate obtained by weight loss measurements (Nishikata et al. 2014). The comb-shaped electrode also enabled the monitoring of time of wetness (TOW). TOW was successfully evaluated as the time period when the solution conductance  $R_s^{-1}$  (from the high-frequency impedance) exceeded a threshold value (e.g.,  $1 \times 10^{-4} \Omega^{-1}$ ) (Nishikata et al. 2005). The TOW measurements were found to depend on the nature of the rusts formed on the electrodes which were made of different types of steels.

While promoting the successful application of EIS as a tool for monitoring the corrosion of metals under thin electrolyte films, Tsuru and Nishikata et al. also realized several drawbacks of their electrodes. First, they discovered that the 10 mHz impedance could not be taken as the polarization resistance when the surface was covered with thick corrosion product (Yadav et al. 2004b). However, a proportional relation still existed between corrosion rate and the reciprocal of the impedance value at 10 mHz. Second, for TOW monitoring, it was found that the electrodes were not completely dry when the measured  $R_s^{-1}$  was zero, possibly due to the presence of discontinuous moisture film which could not conduct across the entire electrode properly.

Other researchers also used similar electrodes to monitor the corrosion of metals under thin electrolyte films. Dong et al. employed the comb-shaped electrodes to study the corrosion of carbon steel and weathering steel under simulated atmospheric conditions (Fu et al. 2009; Thee et al. 2014). The effect of electrolyte drying/thinning on the corrosion rates of both non-corroded electrodes and corroded electrodes (covered with rust) was investigated. They found that the critical electrolyte thickness (i.e., when corrosion rate reached maximum) was relatively larger for the rust-covered electrodes than the mostly pristine electrodes. Li et al. also used the

comb-shaped electrodes to monitor the corrosion rates of weathering steels in a laboratory-simulated environment with varying relative humidity (i.e., 5–100% RH) and verified the EIS monitoring results by weight loss measurements (Li et al. 2010). Huang et al. modified the two-electrode system by introducing a reference electrode through a salt bridge channel fabricated inside the epoxy block, to minimize the ohmic potential drop between the working electrode and reference electrode (Huang et al. 2011). Both cathodic polarization and EIS measurements were conducted to investigate the effects of chloride concentration and relative humidity on the corrosion of printed circuit board (copper) under thin electrolyte layers. The same three-electrode system was also used by Cheng et al. in the investigation of zinc corrosion under thin electrolyte layers (Cheng et al. 2013). Ma et al. used a more sophisticated multielectrode array for monitoring corrosion under thin electrolyte layers (Ma et al. 2015). It was believed that this configuration could provide more accurate corrosion rate measurements under thin and even non-continuous water layers than the two-electrode systems.

## Corrosion in Concrete

Corrosion of reinforcement in concrete is a major reason for the reduced service life of reinforced concrete structures exposed to aggressive environments (e.g., marine). Therefore, monitoring the corrosion of reinforcement is crucial for evaluating the health status of reinforced concrete. Several techniques can be employed to evaluate the corrosion in reinforced concrete, and among the most commonly used methods are the electrochemical techniques. EIS is an attractive technique because it is nearly nondestructive to the structure at the time of application, due to the very small perturbation in the range of 5–20 mV. In addition, EIS is suitable for highly resistive environments such as concrete because it is essentially a transient technique which does not require the system to be in a steady-state condition (John et al. 1981). For steel/concrete systems, information on several parameters such as the presence of surface films, characteristics of concrete, interfacial corrosion, and mass transfer phenomena can be obtained (John et al. 1981). The corrosion rates of the reinforcement can be determined from the  $R_p$  values derived from EIS measurements through the Stern-Geary equation. In addition, the high-frequency impedance provides information about the dielectric properties of concrete, while low-frequency impedance gives information about the characteristics of passive films on steel.

Earlier work conducted three decades ago explored the validity of EIS as a technique for studying the corrosion of rebar in concrete, both experimentally and theoretically. John et al. monitored the corrosion of rebar in concrete with high porosity using EIS (John et al. 1981). They obtained both the corrosion rates of the steel rebar and information about the surface films on the steel. Later, fundamental work was conducted by MacDonald et al. to explore the applicability of EIS for detecting corrosion of rebar in concrete (MacDonald et al. 1988). The rebar was simulated as a 1-D electrical transmission line. Their results suggested that the real and imaginary components of the complex impedance and the phase angle may be

used for detecting corrosion of rebar embedded in concrete, but only at very low frequencies (i.e., <1 mHz). It was also discovered that monitoring of the peak voltage at the concrete surface immediately above the rebar was particularly attractive for corrosion detection. Through EIS measurements, Lay et al. not only determined the corrosion rates of rebar in concrete, but also probed the quality of the concrete and explored the dynamics of wetting, drying, and electrolyte exchange inside the concrete (Lay et al. 1985).

A large amount of work explored the use of low-frequency impedance for monitoring the status of reinforcement in concrete. For example, Montemor et al. investigated the effect of fly ash addition to concrete on the corrosion of reinforcement using EIS (Montemor et al. 1993). EIS measurements were conducted using two steel reinforcement bars with a distance of 6 cm inside the concrete blocks. Detailed analyses on the slopes in the absolute impedance Bode plots were conducted, to obtain information on the state of the steel reinforcement surface. A slope of  $-1$  in the low-frequency region in the Bode plot was observed in a state of passivity, whereas a slope of  $-1/2$  indicated severe corrosion occurring under oxygen diffusion control and a slope of  $-2/3$  was observed under localized attack. Other researchers also reported decreasing slope of the low-frequency impedance with increasing severity of the rebar corrosion caused by increasing chloride concentration (Vedalakshmi et al. 2009). In addition to slope analysis, the low-frequency impedance was widely used to obtain the corrosion rates of reinforcement. Gu et al. used the impedance obtained at the low-frequency range (i.e., 100 Hz–10 mHz) to assess the corrosion of reinforcing steel embedded in cementitious materials containing chloride ions (Gu et al. 1994). Similar work was also conducted by Vedalakshmi et al. (2009).

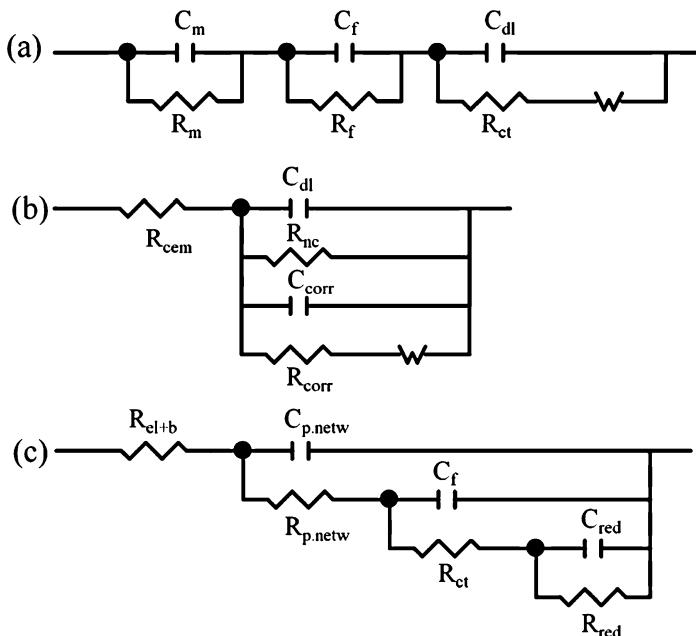
The constant impedance at the high frequencies was interpreted as the ohmic resistance or the resistance of concrete, which was found to increase with the duration of immersion (Scuderi et al. 1991; Christensen et al. 1994; McCarter 1996). However, it was suggested that the high-frequency impedance showed artifacts if a reference electrode was used (Poupard et al. 2003). Poupard et al. claimed that this issue could be resolved by attaching a platinum wire (100  $\mu\text{m}$  in diameter) and a 1000  $\mu\text{F}$  condenser in parallel to the reference electrode.

The above discussion showcases the simple use of low- and high-frequency impedance data for monitoring the corrosion of reinforcement and status of concrete. Extensive efforts were also devoted to fitting the EIS data using different equivalent circuit models. A simple equivalent circuit consisting of the electrolyte resistance ( $R_e$ ) in series with a parallel RC branch has been used to model the steel/concrete system, with R representing the charge transfer resistance ( $R_t$ ) of the corrosion process and C being the double-layer capacitance ( $C_{dl}$ , or the non-ideal constant phase element (CPE)) (Feliu et al. 1998a). When the impedance spectra show more than one semi-circle in the Nyquist plots (i.e., time constants), equivalent circuits with more than one RC combination are necessary. When the impedance spectra exhibit low-frequency tails, a Warburg element ( $W$ ) representing diffusion phenomenon has to be considered.

Figure 3a illustrates an equivalent circuit consisting of three parallel RC components (one with the Warburg component) corresponding to the double-layer,

interfacial film, and the concrete matrix (Sagoe-Crentsil et al. 1992). An intermediate frequency arc in the impedance spectra appeared after certain time periods, which was attributed to the dielectric properties of a surface layer on the rebar. This feature was successfully fitted using the middle RC component in the equivalent circuit shown in Fig. 3a. Similar equivalent circuits were also used by other researchers. Feliu et al. incorporated elements representing diffusion ( $W$ ) and non-ideal capacitive behavior (CPE) to Randles circuit to model the steel/concrete interface (Feliu et al. 1998a). They also explored the accuracy of corrosion rates determined from impedance data (Feliu et al. 1998b).

More complicated equivalent circuits have been proposed to simulate rebar/concrete interface. Gu et al. used another type of equivalent circuit with components simulating the steel surface oxidation behavior (Gu et al. 1994). In their equivalent circuit (Fig. 3b), the  $R_{cem}$  (recorded at frequencies  $>10$  kHz) is the overall bulk resistance consisting of bulk high-frequency resistance and bulk solid/liquid interface resistance. It is related to the porosity, pore size distribution, and pore conductivity. Other components include  $C_{dl}$ , steel/cement interface capacitance;  $R_{nc}$ , surface charge transfer resistance;  $C_{corr}$ , overall capacitance in the corroded region;  $R_{corr}$ , corrosion resistance; and  $W$ , mass transfer resistance (Fig. 3b). Koleva et al. found that EIS was readily applicable for evaluating the electrochemical behavior of steel surfaces under cathodic protection, although the interpretation of derived parameters was not straightforward (Koleva et al. 2007). A relatively



**Fig. 3** Representative equivalent circuits used to simulate the rebar/concrete interfaces (Sagoe-Crentsil et al. 1992; Gu et al. 1994; Koleva et al. 2007)

complicated equivalent circuit model was proposed by taking into consideration all the interfaces between steel/corrosion product/concrete/electrolyte (Fig. 3c). The components  $R_{p,\text{netw}}$  and  $C_{p,\text{netw}}$  are attributed to the properties of the concrete matrix in terms of pore network. The components  $R_{\text{ct}}$  and  $C_f$  are ascribed to the electrochemical reactions on the steel surface, while  $R_{\text{red}}$  and  $C_{\text{red}}$  are related to the redox processes occurring in the corrosion product layers (Koleva et al. 2007).

In addition to the monitoring of rebar corrosion, other applications of EIS on concrete integrity were also investigated. For example, EIS was used to detect moisture inside concrete and the delamination of carbon fiber-reinforced plastics (CFRP). Simple single-frequency measurements were suitable for moisture detection while equivalent circuit modeling was needed to analyze data for CFRP delamination and more reliable moisture determination (Davis et al. 2004; Hong and Harichandran 2005). The EIS technique was also applied to study the effect of tensile and compressive stresses on the degradation of concrete structures (Feng et al. 2016). The results showed that the rebar in compressed concrete samples was more severely corroded than that under the tensile conditions.

## Coatings and Films

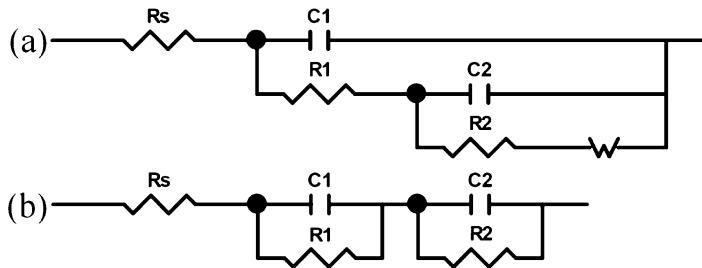
Surface treatment such as the application/formation of protective coatings and films is one of the most important methods to protect metal surfaces. There are various techniques for evaluating coating performance. Among them, EIS is a facile technique which can provide useful information on the protective ability of coatings and their degradation processes, usually through periodic evaluations in accelerated or natural environments. EIS has been considered as a valuable and powerful tool to assess coating degradation, rank coating systems, and predict the service lifetime of coatings, with distinct advantages over DC methods (e.g., linear polarization resistance technique). EIS data are also assimilated to develop meaningful models to analyze the physical behavior of coating degradation. Important parameters such as coating resistance, capacitance, and water uptake can be obtained by interpreting the EIS spectra through equivalent circuit modeling.

The transport of corrosive species (water, oxygen, and detrimental ions) to the coating/metal interface affects the service lifetime of organic coatings. As the corrosive species diffuse into and through coatings, coatings become permeable to water and ions, and eventually, corrosion initiates and propagates at the coating/metal interface. The rate of the interfacial corrosion depends on the transport of corrosive species through the coating that includes the diffusion of water and its accumulation at the coating/metal interface (i.e., water uptake), the diffusion and conduction of ions through the coating, and finally the nature of corrosion reactions.

For the purpose of evaluating coating degradation, EIS is a well-established nondestructive quantitative method to study the water/ions penetration process and the interfacial corrosion. Usually, a three-electrode experimental setup is used for EIS measurements. The electrochemical cell consists of working electrode (i.e., the coated metal), counter electrode, and reference electrode, as mentioned above in the

“Instrumentation” section. For coating measurements, the EIS scanning frequency range is usually from  $10^{-2}$  to  $10^5$  Hz with 10 points per decade using a 10 mV AC perturbation, under the open circuit potential (OCP) condition. The response of the electrochemical cell to the small-voltage perturbation of various frequencies can yield useful information on the properties of the protective coating system. The EIS results are usually presented in the Bode plot as well as the Nyquist plot for coatings. The Bode plot displays the impedance change and/or phase angle change of the coating as a function of frequency. The impedance modulus of EIS spectra at low frequency typically serves as a strong indicator of the corrosion resistance of coatings (Bierwagen et al. 2000; Hinderliter et al. 2008). The semicircles in the Nyquist plot can also illustrate the coating degradation process as shown by the decrease in the semicircle diameter or the appearance of a second semicircle. The Nyquist plot is routinely used to compare with the fitting from equivalent electrical circuit to study coating’s physical behavior during the degradation process.

Equivalent electrical circuit modeling is utilized to interpret EIS data and provide physical explanations for the evaluated phenomena. A simple Randles equivalent circuit is often used to model pristine coatings on metal surfaces (Fig. 2b) (Philippe et al. 2008; Jegdić et al. 2011). When the coatings experience degradation by water and ions penetrating into the coatings, and subsequent corrosion occurs at the metal/coating interface, additional elements have to be considered to represent the physical changes. An equivalent circuit with two RC components in parallel (Fig. 2a) can be used to account for the corrosion occurring under the coating (Murray and Hack 1991, 1992; Scully and Hensley 1994; Armstrong et al. 1995; Gui and Kelly 2005; Macedo et al. 2009; Raps et al. 2009; Singh et al. 2009; Su et al. 2010; Jegdić et al. 2011). This model considers the degradation of coatings with time as water percolates into the coating and a new liquid/metal interface forms under the coating. In this scenario,  $C_1$  is the coating capacitance;  $R_1$  is the pore resistance, which is the resistance of ion-conducting paths that develop in the coating;  $C_2$  is the double-layer capacitance, which is formed as ions from the solution are attracted to the metal surface;  $R_2$  is the charge transfer resistance, which describes the situation of kinetically controlled electrochemical reactions taking place at the metal/solution interface. When water has percolated into the coating, a small area of the coating will delaminate and a pocket filled with solution will form underneath the coating. Thus, the interface between the pocket of solution and the metal substrate can be modeled as a combination of double-layer capacitance in parallel with charge transfer resistance (Zhou and Wang 2013). If the diffusion of corrosive species is the rate-limiting step, a diffusion impedance can be observed in the impedance spectra, and a Warburg element ( $W$ ) can be added to the two RC components (in parallel) circuit (Fig. 4a) (Bretherton et al. 1993; Gui and Kelly 2005). Although most of the EIS data for coatings can be fitted using the equivalent circuits shown in Figs. 2a and 4a, data from some thick coatings with high level of noise cannot be successfully fitted by them. In these cases, a model of two (or more) RC circuits in series may provide better fitting results (Fig. 4b) (Bellucci and Nicodemo 1993). More complicated equivalent circuit models have been proposed for complex coating systems such as coatings with multiple layers (Bretherton et al. 1993; Hernández et al. 2009) and



**Fig. 4** Representative equivalent circuits used to fit EIS data from coating systems (Bretherton et al. 1993; Bellucci and Nicodemo 1993)

composite coatings (Faidi et al. 1993). However, it is still controversial if complicated equivalent circuit models should be used when simple models can generate good fittings.

The following section summarizes some EIS studies on typical coating systems including native oxide films on metals. Since the most prevalent type of coatings used to protect metallic materials is organic coating, this section focuses mainly on EIS evaluation of organic coatings. Specific topics in EIS studies on organic coatings include the investigation of water uptake in the coatings, detection of coating delamination, and in-situ monitoring using embedded sensors. It is a common practice to conduct EIS measurements at certain time intervals when the coated samples are immersed in solutions with varying exposure durations.

Water absorption by coatings on metals is a major factor affecting their protection efficiency. Water diffuses through the polymeric coatings and can cause blister formation, coating delamination, and finally coating failures. Because the relative dielectric constant of polymer is typically very small ( $\epsilon_p = 2-8$ ) compared to that of water ( $\epsilon_w = 80.4$  at  $20^\circ\text{C}$ ), water uptake may cause a large increase of the relative dielectric constant of the coating ( $\epsilon_{coat}$ ). Therefore, water uptake in the coatings significantly changes the coatings' dielectric properties, and the increase of the apparent permittivity of the entire coating systems leads to the increase of the coating capacitance. The research on water uptake in coatings can be traced back to the 1950s (Brasher and Kingsbury 1954) and grew extensively since the 1990s when EIS was developed and applied to coating evaluation studies (Lindqvist 1985; Scully 1989; Walter 1991a, b; van Westing et al. 1994; Bonora et al. 1996; Deflorian et al. 1996; Miskovic-Stankovic et al. 1996; Deflorian et al. 1999; Perez et al. 1999; Castela and Simões 2003a, b, c; Hu et al. 2003; Zhang et al. 2004; Nguyen et al. 2005; Ji et al. 2006; Stafford et al. 2006; Amand et al. 2013).

According to the pioneering work of Brasher and Kingsbury et al., calculation of water uptake of a coating film from observed capacitance change is based on the assumption that the distribution of water into the coating film is random and uniform (Brasher and Kingsbury 1954). The permittivity at time  $t$ ,  $K_t$  is calculated using the following equation:

$$K_t = K_p \times K_w \Phi \quad (9)$$

where  $K_p$  and  $K_w$  are the permittivity of polymer and water, respectively and  $\Phi$  is the volume fraction of absorbed water.

The volume fraction of water can be obtained as follows,

$$\Phi = \frac{\log(C_t/C_0)}{\log \varepsilon_w} \quad (10)$$

where  $C_t$  represents the coating capacitance at time  $t$ ,  $C_0$  represents the coating capacitance at initial immersion, and  $\varepsilon_w$  represents the relative permittivity of water. This equation assumes that the geometrical parameters of the coating are constant over time, i.e., there is no swelling or blistering.

The Brasher and Kingsbury equation has been widely employed to determine the volume fraction of water in coatings, based on the coating capacitance obtained from EIS measurements (Nguyen et al. 2005; Ji et al. 2006). The coating capacitance values are usually obtained by fitting the EIS spectra with appropriate equivalent circuit models. It should be pointed out that, because the properties of coatings evolve during immersion, different models may have to be used at different immersion stages (Hu et al. 2003; Zhang et al. 2004).

Problems with the Brasher and Kingsbury approximation were also realized later (Brasher and Kingsbury 1954). Brasher and Kingsbury attributed the discrepancy to the distribution of pores perpendicular to the surface. Lindqvist compared different formulas for obtaining dielectric properties with the Brasher and Kingsbury equation and concluded that the Brasher and Kingsbury method gave the best agreement with the gravimetric method for small amounts of water uptake (Lindqvist 1985). Popov et al. believed that the coating capacitance change was a measure of water uptake only at the initial stage of exposure (Popov et al. 1993). When the coatings had been exposed for days, coating capacitance change reflected coating delamination and degradation. Stafford et al. found that the Brasher and Kingsbury method was only effective for a random distribution of spherical water inclusions (Stafford et al. 2006). Castela and Simões (Castela and Simões 2003a) found that this approximation gave a good estimation of water volume fraction in clear coatings (Lindqvist 1985) applied on a substrate, but overestimated the water content in commercial coatings (Lindqvist 1985; Walter 1991b; Castela et al. 2000). Therefore, they developed novel impedance models to improve the estimation of water absorption in coatings and their results showed improvements as compared to the classical Brasher and Kingsbury model (Castela and Simões 2003b, c).

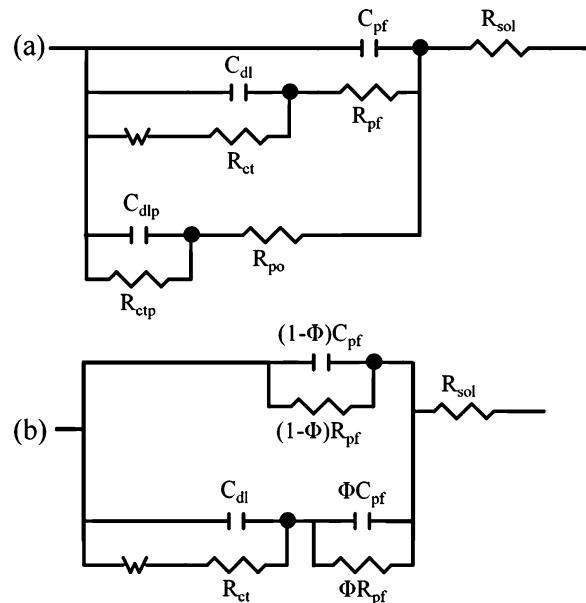
Realizing the discrepancy between the Brasher and Kingsbury method and the gravimetric method, other techniques such as attenuated total reflection infrared spectroscopy (ATR-IR) were employed to investigate the validity of EIS as an effective approach for the study of water uptake (i.e., diffusion coefficient) in organic coatings (Philippe et al. 2008). The three methods Philippe et al. used (i.e., gravimetric, EIS, and ATR-IR) all resulted in two-stage water sorption profiles.

The diffusion coefficients from all three techniques were in close agreement, with those derived from impedance being smaller than others.

In addition to the open circuit condition, the process of water uptake into coatings under cathodic polarization was also studied using EIS (Zhu et al. 2011). Different equivalent circuits were used to fit the EIS spectra obtained from different stages of immersion and the coating capacitance was used to investigate the absorption characteristics of water in coating. It was concluded that water diffusion coefficients were smaller for the coatings under cathodic protection (CP) than those without cathodic protection.

When organic coatings suffer extended exposure to water, coating delamination may occur. This process can be monitored and quantified using EIS (van Westing et al. 1994; Mansfeld 1995). van der Weijde et al. compared the breakpoint frequency (BF) method and the most probable impedance (MPI) equivalent circuit method for the study of coating delamination (van der Weijde et al. 1994). The BF method is based on measuring impedance spectra over a wide range of frequencies, which requires the system to be stable during measurements (Hirayama and Hanuyama 1991; Mansfeld and Tsai 1991). The equivalent circuit to represent the delaminated coatings is shown in Fig. 5a, which has the delamination components and the defect components completely separated. The MPI method uses a different equivalent circuit, shown in Fig. 5b, which has the delamination components in series with a part of the coating impedance (van Westing et al. 1994). Their results indicated that the equivalent circuit in Fig. 5b (MPI) gave better fitting results for delaminated coatings.

**Fig. 5** Relatively complicated equivalent circuit models used to simulate EIS spectra obtained from corrosion under delaminated coating systems (van der Weijde et al. 1994)



Local EIS, or local electrochemical impedance mapping (LEIM), has also been utilized for the study of delamination and/or defects in organic coatings (Wittmann et al. 1999; Mierisch et al. 2003; Jorcin et al. 2006; Foyet et al. 2013; Shkirskiy et al. 2017). It was found that LEIM could detect the formation of blisters in coatings even before they became visible by an optical technique (Wittmann et al. 1999; Mierisch et al. 2003). Foyet et al. conducted local EIS measurements over the defective regions in different coatings on AA2024 substrate and proposed a corrosion protection mechanism based on the spreading of electrons from the defect center to the coatings (Foyet et al. 2013). Shkirskiy et al. successfully used LEIM to quantify the size of defective regions in coatings or the electrochemically active regions (Shkirskiy et al. 2017). They were also able to follow the delamination kinetics of organic coatings on metals.

The above EIS monitoring of coating systems is mostly conducted directly on coated samples. Another approach involves using embedded sensors in coatings to monitor changes in the coatings and corrosion underneath the coatings (Kittel et al. 2001, 2003; Bierwagen et al. 2003; Miszczyk and Schauer 2005; Allahar et al. 2008, 2010; Su et al. 2008). This technique was initially explored by Kittel et al. (Kittel et al. 2001, 2003). They inserted conducting, interdigitated, and thin (1–3 µm) grids made of gold or platinum between two layers of coatings. By doing so, they were able to separate the impedance of the inner part of the coating in contact with the substrate from that of the outer part in contact with the solution. The comb-like electrodes were fabricated by physical vapor deposition (PVD) and served as reference and counter electrodes during EIS measurements. Their results suggested that the embedded electrodes could be used to monitor water uptake in the two layers of a topcoat/primer system simultaneously (Kittel et al. 2001, 2003). The embedded EIS sensors were also used to evaluate the interlayer adhesion of coatings (Miszczyk and Schauer 2005) and monitor the degradation of coatings under various conditions, such as AC-DC-AC accelerated test conditions (Allahar et al. 2008) and thermal cycling (Su et al. 2008). Lastly, Zhou et al. developed a novel three-electrode EIS setup for inline monitoring of coating degradation under flowing conditions (Zhou et al. 2012, 2014). Their unique electrochemical cell consisted of a pair of platinum meshes as the counter electrode and another platinum electrode near the flow outlet as the reference electrode.

Finally, metallic coatings and passive oxide films have also been investigated using the EIS monitoring method. Examples of metallic coatings include zinc coatings on galvanized steel (Hamlaoui et al. 2008) and Ni-W coatings on carbon steel (Shreeram et al. 2017). Examples of passive films include those on stainless steel (Hirschorn et al. 2010; Mohammadi et al. 2011) and aluminum alloys (Mansfeld 1988; Martin et al. 2005).

## Inhibitors

The EIS technique can also be used to study the performance of corrosion inhibitors such as inhibitor film persistency. Considerable research has been conducted to study

corrosion inhibitors for a variety of metals using EIS. This includes mild or carbon steels (Tan et al. 1996; Dhouibi et al. 2002; Gusmano et al. 2006; Khaled and Amin 2009; Li et al. 2011; Popova et al. 2011; Döner et al. 2013), stainless steel (Markhalil et al. 2013), copper alloys (Lokesh et al. 2012), aluminum alloys (Sherif and Park 2005, 2006; Zheludkevich et al. 2005), and magnesium alloys (Gao et al. 2010).

For the extensively studied systems of mild steels with inhibitors, the EIS spectra obtained can be fitted using a simple, one RC component equivalent circuit similar to the one shown in Fig. 2b (Li et al. 2011; Döner et al. 2013; Sığircik et al. 2017). From the EIS spectra fitting, the solution resistance ( $R_s$ ), charge transfer resistance ( $R_{ct}$ ), and double-layer capacitance ( $C_{dl}$ ) are obtained. The charge transfer resistance can then be used to calculate inhibition efficiency (%I) by the following expression:

$$(\%) I = \frac{|R_{ct} - R_{ct}^0|}{R_{ct}} \times 100 \quad (11)$$

where  $R_{ct}$  and  $R_{ct}^0$  are the charge transfer resistance of the metal with and without inhibitors, respectively.

Tan et al. used EIS to study the mechanism of inhibitor film formation and destruction and evaluate the film persistency of a commercial CO<sub>2</sub> corrosion inhibitor: imidazoline (Tan et al. 1996). Parameters such as the resistance and capacitance of different inhibitor layers, charge transfer resistance, and double-layer capacitance were obtained for analyzing the inhibition mechanism and calculating corrosion rates. Two types of equivalent circuits were proposed by Tan et al., which had additional RC elements for all four layers of non-conducting inhibitor films (Tan et al. 1996). Similarly, two imidazoline-based inhibitors for the protection of mild steels in high-salinity and CO<sub>2</sub>-saturated solutions were characterized using EIS (Gusmano et al. 2006). Different mechanisms for inhibition performance were identified for inhibitors with different formulations, and therefore different equivalent circuit models were used to fit the impedance data (Gusmano et al. 2006). Additional elements were added to the equivalent circuits for modeling the impedance data from the inhibitor that resulted in a secondary absorbed layer.

In addition to the equivalent circuits composed of only RC components, more complicated models with other elements, such as inductance ( $L$ ) and Warburg resistance ( $W$ ), were used to study inhibitor systems. Khaled et al. investigated the inhibition performance of three thiazole derivatives for mild steel in 0.5 M H<sub>2</sub>SO<sub>4</sub> (Khaled and Amin 2009). The EIS spectra of these corrosion inhibition systems exhibited an inductive loop in the low-frequency range, in addition to the commonly observed capacitive loop in the high-frequency range. The depressed high-frequency loop indicated surface inhomogeneity, similar to those found in adsorption processes (Gonçalves et al. 2002). The formation of an inductive loop was attributed to the relaxation process caused by adsorbed species (e.g.,  $H_{ads}^+$ ) on the metal surface and/or redissolution of the passivated surface (Sherif and Park 2006). In other inhibitor systems, such as those for copper (Lokesh et al. 2012; Mendonça et al. 2017), Warburg impedance ( $W$ ) was observed in the EIS spectra, and therefore Warburg

component was added to the equivalent circuits. The Warburg impedance was related to the diffusion processes and suggested the porous nature of the inhibitor films (Lokesh et al. 2012).

## High-Temperature Applications

High-temperature corrosion occurs under conditions such as high-temperature steam, high-temperature water, and molten salts. EIS has been explored as an effective technique to monitor hot corrosion under these conditions.

High-temperature water conditions include, for example, high-temperature and high-pressure steam (Durand-keklikian et al. 1991), boiling water reactors (BWR) (Asakura et al. 1989; Kim 2000; Yeh et al. 2008), pressurized water reactors (PWR) (Park and Macdonald 1983; Macák et al. 2006; Bojinov et al. 2010, 2014; Betova et al. 2012; Zebardast et al. 2013), and supercritical water reactors (SCWR) (Betova et al. 2006, 2007; Fulger et al. 2009, 2011). The experiments in these conditions can be conducted using the conventional three-electrode system but with special reference electrode designed for high-temperature applications. The materials frequently studied in high-temperature water conditions include stainless steels, Inconel alloys, nickel alloys, and zirconium alloys. For example, Kim suggested that stainless steel surfaces were usually covered with a relatively homogeneous oxide layer in high-temperature water conditions, and therefore, the simple Randles equivalent circuit was suitable for fitting the EIS data (Kim 2000). Fulger et al. simply obtained the well-defined  $|Z|$  values from the low-frequency end and used them as an indication of corrosion resistance (Fulger et al. 2009). Others proposed complicated equivalent circuit models (Macák et al. 2006) and physical models (Betova et al. 2007, 2012; Bojinov et al. 2010) to interpret the EIS data. For example, Bojinov et al. used a transfer function derived from the mixed conduction model to fit the EIS data and obtained kinetic parameters for the oxidation process, including interfacial rate constant of oxidation, diffusion coefficient of oxygen vacancies, and field strength in the inner oxide layer (Bojinov et al. 2010). Besides the characteristics and mechanism of high-temperature oxidation, corrosion product fouling process in high-temperature and high-pressure solution was also successfully monitored using the EIS technique (Zebardast et al. 2013).

The corrosion of metals in contact with molten salts also occurs through electrochemical mechanism and EIS is one of the electrochemical techniques that have been proven to be effective in investigating the reaction mechanisms under molten salts. The metal degradation in molten salts environment has been investigated either in fully-immersed conditions (Perez et al. 1999; Yang and Kim 1999; Zhu et al. 1999; Zeng et al. 2000, a, b; Takeuchi et al. 2005; Frangini and Loretí 2006) or under thin films of molten salts (Gao et al. 1990; Wu and Rapp 1991; Zeng and Li 2005; Pan et al. 2009; Ni et al. 2011). These studies were usually conducted in molten sulfates (e.g.,  $\text{Na}_2\text{SO}_4$ , pure or with the addition of  $\text{NaCl}$  and/or  $\text{K}_2\text{SO}_4$ ), carbonates (e.g.,  $\text{K}_2\text{CO}_3$ - $\text{Li}_2\text{CO}_3$ ), and chlorides (e.g.,  $\text{ZnCl}_2$ - $\text{KCl}$ ). In such conditions, EIS measurements were typically conducted using a two-electrode system. The two electrodes were commonly made of the same test materials, with one electrode serving as the

working electrode and the other acting as both counter electrode and reference electrode. Different equivalent circuit models were selected for different conditions. For example, Zeng et al. proposed four equivalent circuit models for electrodes Pt, Ni<sub>3</sub>Al, and FeAl in molten salt systems (Zeng et al. 2001). Charge transfer process was found to be the rate-limiting step when protective scales formed on the metal surfaces, while characteristics of diffusion-controlled reactions were observed when the scales were non-protective (Zeng et al. 2001).

Aung et al. developed a high-temperature electrochemical sensor based on the EIS technique and successfully applied it to monitor hot corrosion, such as corrosion in coal-fired boiler environments (Aung and Liu 2012, 2013, 2014). They used a four-electrode system (Fig. 1b) with two identical working electrodes made of the test materials. The reference electrode was an Ag/Ag<sup>+</sup>/fused-quartz electrode containing a silver electrode wire immersed in sulfate electrolyte. A platinum wire served as the counter electrode. They were able to monitor reaction kinetics in hot corrosion process using this sensor. For example, it was found that at T = 750 or 800 °C, the corrosion process on Inconel 740 alloy was charge transfer-controlled in the first 49.5 h, followed by the diffusion-controlled process. At T = 700 °C, only charge transfer-controlled process was observed, possibly due to the delayed formation of low-melting alkali sulfate eutectic mixtures.

## Other Applications

The EIS technique has also been explored for corrosion monitoring in other areas, in addition to the abovementioned topics. This section gives a brief review on the use of EIS for corrosion monitoring in oil and gas-related applications, microbiologically influenced corrosion (MIC), and stress corrosion cracking (SCC). The possible limitations of EIS measurements in specific applications are also pointed out.

In the oil and gas industry, EIS has been used to study various corrosion phenomena, such as: (1) CO<sub>2</sub> corrosion of mild steels in brine (i.e., NaCl) environments (Kinsella et al. 1998; Bian et al. 2015; Zhu et al. 2015; Zeng et al. 2016), and (2) soil corrosion of buried pipeline steels (Murray and Moran 1989; Choi and Kim 2005; Choi et al. 2006, 2007). In CO<sub>2</sub>-containing brine environments, EIS was used to continuously monitor the formation and growth of corrosion product films (mostly carbonates) on carbon steels. Different equivalent circuit models were chosen to fit the EIS data at different stages due to the fact that the corrosion product film was non-uniform in the initial stage, but a full surface coverage was achieved after longer exposure (Zhu et al. 2015). The EIS investigations on soil corrosion are similar to those in bulk solutions because aqueous corrosion occurs when the metals are in contact with soil solutions. Correspondingly, the equivalent circuit models for fitting EIS spectra are also similar. For example, Choi et al. used the simple equivalent circuit model shown in Fig. 2b to fit the EIS spectra obtained in the initial stage of immersion when there was no rust layer on the metals. After the formation of a rust layer, two time constants were observed from the EIS spectra, and therefore, the equivalent circuit model shown in Fig. 2a was proper to use (Choi and Kim 2005).

The process of MIC has been monitored using the EIS technique (Dowling et al. 1988; Mansfeld and Little 1991; Liu et al. 2016; Zhao et al. 2017). The occurrence of MIC is due to the presence of microorganisms on metal surfaces, which usually form biofilms. Since MIC is a bio-electrochemical process, electrochemical measurements such as EIS can be applied to achieve real-time monitoring of the MIC process. However, it was recently suggested that continuous EIS measurements could have inhibited cell growth and reduced biofilm thickness (Zhao et al. 2017). As a result, inaccurate corrosion rate and corrosion pit depth information could have been obtained from EIS analyses. Complementary techniques such as electrochemical noise (EN) should be utilized (Zhao et al. 2017).

Finally, SCC can also be investigated using EIS (Oltra and Keddamt 1988; Petit et al. 1990; Bosch et al. 2001; Bosch 2005; Oskuiie et al. 2012; Carpintero-Moreno et al. 2013), although it is commonly believed that the presence of local events during SCC would produce a lot of data dispersion. Petit et al. observed phase shifts at particular frequencies in the EIS spectra obtained during SCC tests and correlated them to crack initiations (Petit et al. 1990). Bosch et al. developed a crack model for the impedance data obtained during SCC and concluded that the phase shifts at particular frequencies were related to the length of the cracks (Bosch 2005). In addition, it was suggested that the phase shifts might be too small due to sensitivity issues in some systems.

---

## Summary

Among the various electrochemical techniques used for corrosion monitoring, EIS is believed to be one of the least destructive, because the amplitudes of the currents and voltages involved in the measurements are typically very small. The AC characteristic of the technique makes it suitable for highly resistive systems. These advantages lead to the broad and successful applications of EIS as a powerful electrochemical technique for corrosion monitoring. The topics covered in this chapter include atmospheric corrosion, corrosion in concrete, coatings and films, inhibitors, high-temperature corrosion, oil and gas industry-related corrosion, microbiologically influenced corrosion, and stress corrosion cracking. The applicability of various electrode systems in different conditions is also discussed. Finally, it should be emphasized that although EIS measurement is relatively straightforward, EIS data analysis/fitting is often complicated. Therefore, extra caution should be taken when selecting the appropriate equivalent circuit models for data analysis, and mathematical models that are based on actual physical properties of the systems should be proposed.

---

## References

- Allahar KN, Su Q, Bierwagen GP, Lee D (2008) Monitoring of the AC-DC-AC degradation of organic coatings using embedded electrodes. *Corrosion* 64:773–787  
Allahar K, Su Q, Bierwagen G (2010) Non-substrate EIS monitoring of organic coatings with embedded electrodes. *Prog Org Coat* 67:180–187

- Amand S, Musiani M, Orazem ME et al (2013) Constant-phase-element behavior caused by inhomogeneous water uptake in anti-corrosion coatings. *Electrochim Acta* 87:693–700
- Armstrong RD, Jenkins ATA, Johnson BW (1995) An investigation into the UV breakdown of thermoset polyester coatings using impedance spectroscopy. *Corros Sci* 37:1615–1625
- Asakura Y, Karasawa H, Sakagami M, Uchida S (1989) Relationships between corrosion behavior of AISI 304 stainless steel in high-temperature pure water and its oxide film structures. *Corrosion* 45:119–124
- Aung NN, Liu X (2012) High temperature electrochemical sensor for in situ monitoring of hot corrosion. *Corros Sci* 65:1–4
- Aung NN, Liu X (2013) Effect of SO<sub>2</sub> in flue gas on coal ash hot corrosion of Inconel 740 alloy – a high temperature electrochemical sensor study. *Corros Sci* 76:390–402
- Aung NN, Liu X (2014) Effect of temperature on coal ash hot corrosion resistance of Inconel 740 superalloy. *Corros Sci* 82:227–238
- Beaumier L, Epelboin I, Lestrade JC, Takenouti H (1976) Etude electrochimique, et par microscopie electronique a balayage, du fer recouvert de peinture. *Surf Technol* 4:237–254
- Bellucci F, Nicodemo L (1993) Water transport in organic coatings. *Corrosion* 49:235–247
- Betova I, Bojinov M, Kinnunen P et al (2006) Composition, structure, and properties of corrosion layers on ferritic and austenitic steels in ultrasupercritical water. *J Electrochem Soc* 153: B464–B473
- Betova I, Bojinov M, Kinnunen P et al (2007) Surface film electrochemistry of austenitic stainless steel and its main constituents in supercritical water. *J Supercrit Fluids* 43:333–340
- Betova I, Bojinov M, Karastoyanov V et al (2012) Effect of water chemistry on the oxide film on alloy 690 during simulated hot functional testing of a pressurised water reactor. *Corros Sci* 58:20–32
- Bian C, Wang ZM, Han X et al (2015) Electrochemical response of mild steel in ferrous ion enriched and CO<sub>2</sub> saturated solutions. *Corros Sci* 96:42–51
- Bierwagen GP, He L, Li J et al (2000) Studies of a new accelerated evaluation method for coating corrosion resistance – thermal cycling testing. *Prog Org Coat* 39:67–78
- Bierwagen GP, Wang X, Tallman DE (2003) In situ study of coatings using embedded electrodes for ENM measurements. *Prog Org Coat* 46:163–175
- Bojinov M, Karastoyanov V, Kinnunen P, Saario T (2010) Influence of water chemistry on the corrosion mechanism of a zirconium-niobium alloy in simulated light water reactor coolant conditions. *Corros Sci* 52:54–67
- Bojinov M, Nowak E, Saario T et al (2014) Effect of chloride on the oxides on low-alloyed steel in conditions of a light water reactor pressure vessel cladding flaw. *J Electrochem Soc* 161: C177–C187
- Bonora PL, Deflorian F, Fedrizzi L (1996) Electrochemical impedance spectroscopy as a tool for investigating underpaint corrosion. *Electrochim Acta* 41:1073–1082
- Bosch RW (2005) Electrochemical impedance spectroscopy for the detection of stress corrosion cracks in aqueous corrosion systems at ambient and high temperature. *Corros Sci* 47:125–143
- Bosch RW, Moens F, Zheng JH, Bogaerts WF (2001) Application of electrochemical impedance spectroscopy for monitoring stress corrosion cracking. *Corrosion* 57:532–539
- Brasher DM, Kingsbury AH (1954) Electrical measurements in the study of immersed paint coatings on metal. I. Comparison between capacitance and gravimetric methods of estimating water-uptake. *J Appl Chem* 4:62–72
- Bretherton N, Thompson GE, Turgoose S (1993) An electrochemical investigation of the development of phosphate conversion coatings. *Corros Sci* 35:1297–1303
- Carpintero-Moreno EJ, Gonzalez-Rodriguez JG, Uruchurtu-Chavarín J et al (2013) Detection of stress corrosion cracking in a dual-phase steel using electrochemical noise and impedance spectroscopy. *Corrosion* 69:85–94
- Castela AS, Simões AM (2003a) Water sorption in freestanding PVC films by capacitance measurements. *Prog Org Coat* 46:130–134

- Castela AS, Simões AM (2003b) An impedance model for the estimation of water absorption in organic coatings. Part II: a complex equation of mixture. *Corros Sci* 45:1647–1660
- Castela AS, Simões AM (2003c) An impedance model for the estimation of water absorption in organic coatings. Part I: a linear dielectric mixture equation. *Corros Sci* 45:1631–1646
- Castela ASL, Simões AM, Ferreira MGS (2000) E.I.S. Evaluation of attached and free polymer films. *Prog Mater Sci* 38:1–7
- Cheng Q, Song S, Song L, Hou B (2013) Effect of relative humidity on the initial atmospheric corrosion behavior of zinc during drying. *J Electrochem Soc* 160:C380–C389
- Choi Y-S, Kim J-G (2005) A galvanic sensor for monitoring the corrosion damage of buried pipelines: part 1 – electrochemical tests to determine the correlation of probe current to actual corrosion damage in synthetic groundwater. *Corrosion* 61:293–301
- Choi Y-S, Kim J-G, Yang SJ (2006) A galvanic sensor for monitoring the corrosion damage of buried pipelines: part 2 – correlation of sensor output to actual corrosion damage of pipeline in soil and tap water environments. *Corrosion* 62:522–532
- Choi Y-S, Kim J-G, Koo J-Y (2007) A galvanic sensor for monitoring the corrosion damage of buried pipelines: part 3 – correlation of probe current to cathodic protection and stray current. *Corrosion* 63:951–957
- Christensen BJ, Coverdale RT, Olson RA et al (1994) Impedance spectroscopy of hydrating cement-based materials: measurement, interpretation, and application. *J Am Ceramic Soc* 77:2789–2804
- Cooper K, Smith M, Budiansky N, Scully J (2006) Development of a multi-electrode array impedance analyzer for corrosion science and sensors. CORROSION/2006 Conference, NACE International, Paper no. 00674
- Davis GD, Rich MJ, Drzal LT (2004) Monitoring moisture uptake and delamination in CFRP reinforced concrete structures with electrochemical impedance sensors. *J Nondestruct Eval* 23:1–9
- de Levie R (1963) On porous electrodes in electrolyte solutions. *Electrochim Acta* 8:751–780
- Deflorian F, Fedrizzi L, Bonora PL (1996) Influence of the photo-oxidative degradation on the water barrier and corrosion protection properties of polyester paints. *Corros Sci* 38:1697–1708
- Deflorian F, Fedrizzi L, Rossi S, Bonora PL (1999) Organic coating capacitance measurement by EIS: ideal and actual trends. *Electrochim Acta* 44:4243–4249
- Dhouibi L, Triki E, Raharinaivo A (2002) The application of electrochemical impedance spectroscopy to determine the long-term effectiveness of corrosion inhibitors for steel in concrete. *Cem Concr Compos* 24:35–43
- Döner A, Şahin EA, Kardaş G, Serindağ O (2013) Investigation of corrosion inhibition effect of 3-[2-hydroxy-benzylidene]-amino]-2-thioxo-thiazolidin-4-one on corrosion of mild steel in the acidic medium. *Corros Sci* 66:278–284
- Dowling NJE, Guezennec J, Lemoine ML et al (1988) Analysis of carbon steels affected by bacteria using electrochemical impedance and direct current techniques. *Corrosion* 44:869–874
- Durand-keklikian L, Cagnolino G, Macdonald DD (1991) Technical note: in situ A.C. Impedance studies of oxide film growth on zircaloy-4 in high temperature/high pressure steam. *Corros Sci* 32:361–365
- El-Mahdy GA, Kim KB (2004) AC impedance study on the atmospheric corrosion of aluminum under periodic wet-dry conditions. *Electrochim Acta* 49:1937–1948
- El-Mahdy GA, Nishikata A, Tsuru T (2000a) AC impedance study on corrosion of 55%Al-Zn alloy-coated steel under thin electrolyte layers. *Corros Sci* 42:1509–1521
- El-Mahdy GA, Nishikata A, Tsuru T (2000b) Electrochemical corrosion monitoring of galvanized steel under cyclic wet-dry conditions. *Corros Sci* 42:183–194
- Faidi SE, Scantlebury JD, Bullivant P et al (1993) An electrochemical study of zinc-containing epoxy coatings on mild steel. *Corros Sci* 35:1319–1328
- Feliu V, González JA, Andrade C, Feliu S (1998a) Equivalent circuit for modelling the steel-concrete interface. I. Experimental evidence and theoretical predictions. *Corros Sci* 40:975–993

- Feliu V, González JA, Andrade C, Feliu S (1998b) Equivalent circuit for modelling the steel-concrete interface. II. Complications in applying the Stern-Geary equation to corrosion rate determinations. *Corros Sci* 40:995–1006
- Feng X, Lu X, Zuo Y et al (2016) Electrochemical study the corrosion behaviour of carbon steel in mortars under compressive and tensile stresses. *Corros Sci* 103:66–74
- Foyet A, Wu TH, Kodentsov A et al (2013) Corrosion protection and delamination mechanism of epoxy/carbon black nanocomposite coating on AA2024-T3. *J Electrochem Soc* 160: C159–C167
- Frangini S, Loreti S (2006) The role of temperature on the corrosion and passivation of type 310S stainless steel in eutectic (Li + K) carbonate melt. *J Power Sources* 160:800–804
- Fu X, Dong J, Han E, Ke W (2009) A new experimental method for in situ corrosion monitoring under alternate wet-dry conditions. *Sensors* 9:10400–10410
- Fulger M, Ohai D, Mihalache M et al (2009) Oxidation behavior of Incoloy 800 under simulated supercritical water conditions. *J Nucl Mater* 385:288–293
- Fulger M, Mihalache M, Ohai D et al (2011) Analyses of oxide films grown on AISI 304L stainless steel and Incoloy 800HT exposed to supercritical water environment. *J Nucl Mater* 415:147–157
- Gao G, Stott FH, Dawson JL, Farrel DM (1990) Electrochemical monitoring of high-temperature molten-salt corrosion. *Oxid Met* 33:79–94
- Gao H, Li Q, Dai Y et al (2010) High efficiency corrosion inhibitor 8-hydroxyquinoline and its synergistic effect with sodium dodecylbenzenesulphonate on AZ91D magnesium alloy. *Corros Sci* 52:1603–1609
- Gonçalves RS, Azambuja DS, Serpa Lúcho AM (2002) Electrochemical studies of propargyl alcohol as corrosion inhibitor for nickel, copper, and copper/nickel (55/45) alloy. *Corros Sci* 44:467–479
- Gu P, Fu Y, Xie P, Beaudoin JJ (1994) Characterization of surface corrosion of reinforcing steel in cement paste by low frequency impedance spectroscopy. *Cem Concr Res* 24:231–242
- Gui F, Kelly RG (2005) Performance assessment and prediction of corrosion prevention compounds with electrochemical impedance spectroscopy. *Corros Sci* 61:119–129
- Gusmano G, Labella P, Montesperelli G et al (2006) Study of the inhibition mechanism of imidazolines by electrochemical impedance spectroscopy. *Corrosion* 62:576–583
- Hamlaoui Y, Pedraza F, Tifouti L (2008) Corrosion monitoring of galvanised coatings through electrochemical impedance spectroscopy. *Corros Sci* 50:1558–1566
- Hernández M, Genescá J, Uruchurtu J, Barba A (2009) Correlation between electrochemical impedance and noise measurements of waterborne coatings. *Corros Sci* 51:499–510
- Hinderliter BR, Allahar KN, Bierwagen GP et al (2008) Water sorption and diffusional properties of a cured epoxy resin measured using alternating ionic liquids/aqueous electrolytes in electrochemical impedance spectroscopy. *J Coat Technol Res* 5:431–438
- Hirayama R, Hanuyama S (1991) Electrochemical impedance for degraded coated steel having pores. *Corrosion* 47:952–958
- Hirschorn B, Orazem ME, Tribollet B et al (2010) Determination of effective capacitance and film thickness from constant-phase-element parameters. *Electrochim Acta* 55:6218–6227
- Hong S, Harichandran RS (2005) Nondestructive evaluation of CFRP/concrete bond in beams using electrochemical impedance spectroscopy. *J Compos Constr* 9:515–523
- Hu JM, Zhang JQ, Cao CN (2003) Determination of water uptake and diffusion of Cl-ion in epoxy primer on aluminum alloys in NaCl solution by electrochemical impedance spectroscopy. *Prog Org Coat* 46:273–279
- Huang H, Dong Z, Chen Z, Guo X (2011) The effects of Cl-ion concentration and relative humidity on atmospheric corrosion behaviour of PCB-Cu under adsorbed thin electrolyte layer. *Corros Sci* 53:1230–1236
- Jegdić BV, Bajat JB, Popić JP et al (2011) The EIS investigation of powder polyester coatings on phosphated low carbon steel: the effect of NaNO<sub>2</sub> in the phosphating bath. *Corros Sci* 53:2872–2880

- Ji WG, Hu JM, Zhang JQ, Cao CN (2006) Reducing the water absorption in epoxy coatings by silane monomer incorporation. *Corros Sci* 48:3731–3739
- John DG, Searson PC, Dawson JL (1981) Use of AC impedance technique in studies on steel in concrete in immersed conditions. *Br Corros J* 16:102–106
- Jorcin JB, Aragon E, Merlatti C, Pébère N (2006) Delaminated areas beneath organic coating: a local electrochemical impedance approach. *Corros Sci* 48:1779–1790
- Khaled KF, Amin MA (2009) Corrosion monitoring of mild steel in sulphuric acid solutions in presence of some thiazole derivatives – molecular dynamics, chemical and electrochemical studies. *Corros Sci* 51:1964–1975
- Kim Y-J (2000) In-situ electrochemical impedance measurement of oxide film formed on type 304 stainless steel in high-temperature water. *Corrosion* 56:389–394
- Kinsella B, Tan YJ, Bailey S (1998) Electrochemical impedance spectroscopy and surface characterization techniques to study carbon dioxide corrosion product scales. *Corrosion* 54:835–842
- Kittel J, Celati N, Keddam M, Takenouti H (2001) New methods for the study of organic coatings by EIS: new insights into attached and free films. *Prog Org Coat* 41:93–98
- Kittel J, Celati N, Keddam M, Takenouti H (2003) Influence of the coating-substrate interactions on the corrosion protection: characterisation by impedance spectroscopy of the inner and outer parts of a coating. *Prog Org Coat* 46:135–147
- Koleva DA, de Wit JHW, van Breugel K et al (2007) Investigation of corrosion and cathodic protection in reinforced concrete I. Application of electrochemical techniques. *J Electrochem Soc* 154:P52–P61
- Lay P, Lawrence PF, Wilkins JM, Williams DE (1985) An A.C. Impedance study of steel in concrete. *J Appl Electrochem* 15:755–766
- Li C, Ma Y, Li Y, Wang F (2010) EIS monitoring study of atmospheric corrosion under variable relative humidity. *Corros Sci* 52:3677–3686
- Li X, Deng S, Fu H (2011) Benzyltrimethylammonium iodide as a corrosion inhibitor for steel in phosphoric acid produced by dihydrate wet method process. *Corros Sci* 53:664–670
- Lindqvist SA (1985) Theory of dielectric properties of heterogeneous substances applied to water in a paint film. *Corrosion* 41:69–75
- Liu H, Gu T, Zhang G et al (2016) The effect of magneticfield on biomineralization and corrosion behavior of carbon steel induced by iron-oxidizing bacteria. *Corros Sci* 102:93–102
- Lokesh KS, De Keersmaecker M, Elia A et al (2012) Adsorption of cobalt (II) 5,10,15,20-tetrakis (2-aminophenyl)-porphyrin onto copper substrates: characterization and impedance studies for corrosion inhibition. *Corros Sci* 62:73–82
- Ma X, Cheng Q, Zheng M et al (2015) Monitoring marine atmospheric corrosion by electrochemical impedance spectroscopy under various relative humidities. *Int J Electrochem Sci* 10:10402–10421
- Macák J, Sajdl P, Kučera P et al (2006) In situ electrochemical impedance and noise measurements of corroding stainless steel in high temperature water. *Electrochim Acta* 51:3566–3577
- Macdonald JR (1987) Impedance spectroscopy. Wiley, New York
- MacDonald DD, McKubre MCH, Urquidi-MacDonald M (1988) Theoretical assessment of AC impedance spectroscopy for detecting corrosion of rebar in reinforced concrete. *Corrosion* 44:2–7
- Macedo MCSS, Margarit-Mattos ICP, Fragata FL et al (2009) Contribution to a better understanding of different behaviour patterns observed with organic coatings evaluated by electrochemical impedance spectroscopy. *Corros Sci* 51:1322–1327
- Mansfeld F (1988) Evaluation of anodized aluminum surfaces with electrochemical impedance spectroscopy. *J Electrochem Soc* 135:828–833
- Mansfeld F (1995) Use of electrochemical impedance spectroscopy for the study of corrosion protection by polymer coatings. *J Appl Electrochem* 25:187–202
- Mansfeld F, Little B (1991) A technical review of electrochemical techniques applied to microbiologically influenced corrosion. *Corros Sci* 32:247–272

- Mansfeld F, Tsai CH (1991) Determination of coating deterioration with EIS. I. Basic relationships. *Corrosion* 47:958–963
- Markhali BP, Naderi R, Mahdavian M et al (2013) Electrochemical impedance spectroscopy and electrochemical noise measurements as tools to evaluate corrosion inhibition of azole compounds on stainless steel in acidic media. *Corros Sci* 75:269–279
- Martin FJ, Cheek GT, O'Grady WE, Natishan PM (2005) Impedance studies of the passive film on aluminium. *Corros Sci* 47:3187–3201
- McCarter WJ (1996) The A.C. Impedance response of concrete during early hydration. *J Mater Sci* 31:6285–6292
- Mendonça GLF, Costa SN, Freire VN et al (2017) Understanding the corrosion inhibition of carbon steel and copper in sulphuric acid medium by amino acids using electrochemical techniques allied to molecular modelling methods. *Corros Sci* 115:41–55
- Mierisch AM, Taylor SR, Celli V (2003) Understanding the degradation of organic coatings through local electrochemical impedance methods. *J Electrochem Soc* 150:B309–B315
- Miskovic-Stankovic VB, Dražić DM, Kačarević-Popović Z (1996) The sorption characteristics of epoxy coatings electrodeposited on steel during exposure to different corrosive agents. *Corros Sci* 38:1513–1523
- Miszczyk A, Schauer T (2005) Electrochemical approach to evaluate the interlayer adhesion of organic coatings. *Prog Org Coat* 52:298–305
- Mohammadi F, Nickchi T, Attar MM, Alfantazi A (2011) EIS study of potentiostatically formed passive film on 304 stainless steel. *Electrochim Acta* 56:8727–8733
- Montemor MF, Simoes AMP, Salta MM, Ferreira MGS (1993) The assessment of the electrochemical behaviour of flyash-containing concrete by impedance spectroscopy. *Corros Sci* 35:1571–1578
- Murray JN, Hack HP (1991) Long-term testing of epoxy-coated steel in ASTM seawater using electrochemical impedance spectroscopy. *Corrosion* 47:480–489
- Murray JN, Hack HP (1992) Testing organic architectural coatings in ASTM synthetic seawater immersion conditions using EIS. *Corrosion* 48:671–685
- Murray JN, Moran PJ (1989) Influence of moisture on corrosion of pipeline steel in soils using in situ impedance spectroscopy. *Corrosion* 45:34–43
- Nguyen VN, Perrin FX, Vernet JL (2005) Water permeability of organic/inorganic hybrid coatings prepared by sol-gel method: a comparison between gravimetric and capacitance measurements and evaluation of non-Fickian sorption models. *Corros Sci* 47:397–412
- Ni CS, Lu LY, Zeng CL, Niu Y (2011) Electrochemical impedance studies of the initial-stage corrosion of 310S stainless steel beneath thin film of molten  $(0.62\text{Li}, 0.38\text{K})_2\text{CO}_3$  at 650 °C. *Corros Sci* 53:1018–1024
- Nishikata A, Ichihara Y, Tsuru T (1995a) An application of electrochemical impedance spectroscopy to atmospheric corrosion study. *Corros Sci* 37:897–911
- Nishikata A, Yamashita Y, Isatayama H et al (1995b) An electrochemical impedance study on atmospheric corrosion of steels in a cyclic wet-dry condition. *Corros Sci* 37:2059–2069
- Nishikata A, Ichihara Y, Hayashi V, Tsuru T (1997) Influence of electrolyte layer thickness and pH on the initial stage of the atmospheric corrosion of iron. *J Electrochem Soc* 144:1244–1252
- Nishikata A, Suzuki F, Tsuru T (2005) Corrosion monitoring of nickel-containing steels in marine atmospheric environment. *Corros Sci* 47:2578–2588
- Nishikata A, Zhu Q, Tada E (2014) Long-term monitoring of atmospheric corrosion at weathering steel bridges by an electrochemical impedance method. *Corros Sci* 87:80–88
- Oltra R, Keddamt M (1988) Application of impedance technique to localized corrosion. *Corros Sci* 28:1–18
- Oskuiie AA, Shahabi T, Shahriari A, Saebnoori E (2012) Electrochemical impedance spectroscopy analysis of X70 pipeline steel stress corrosion cracking in high pH carbonate solution. *Corros Sci* 61:111–122
- Pan TJ, Lu WM, Ren YJ et al (2009) Electrochemical-impedance-spectroscopy (EIS) study of corrosion of steels 12CrMoV and SS304 beneath a molten  $\text{ZnCl}_2\text{-KCl}$  film at 400 °C in air. *Oxid Met* 72:179–190

- Park JR, Macdonald DD (1983) Impedance studies of the growth of porous magnetite films on carbon steel in high temperature aqueous systems. *Corros Sci* 23:295–315
- Perez C, Collazo A, Izquierdo M et al (1999) Characterisation of the barrier properties of different paint systems. Part I. Experimental set-up and ideal Fickian diffusion. *Prog Org Coat* 36:102–108
- Petit MC, Cid M, Puiggali M, Amor Z (1990) An impedance study of the passivity breakdown during stress corrosion cracking phenomena. *Corros Sci* 31:491–496
- Philippe LVS, Lyon SB, Sammon C, Yarwood J (2008) Validation of electrochemical impedance measurements for water sorption into epoxy coatings using gravimetry and infra-red spectroscopy. *Corros Sci* 50:887–896
- Popov BN, Alwohaibi MA, White RE (1993) Using electrochemical impedance spectroscopy as a tool for organic coating solute saturation monitoring. *J Electrochem Soc* 140:947–951
- Popova A, Christov M, Vasilev A (2011) Mono- and dicationic benzothiazolic quaternary ammonium bromides as mild steel corrosion inhibitors. Part II: electrochemical impedance and polarisation resistance results. *Corros Sci* 53:1770–1777
- Poupard O, Ait-Mokhtar A, Dumargue P (2003) Impedance spectroscopy in reinforced concrete: experimental procedure for monitoring steel. Part I development of the experimental device. *J Mater Sci* 38:2845–2850
- Rammelt U, Reinhard G (1992) Application of electrochemical impedance spectroscopy (EIS) for characterizing the corrosion-protective performance of organic coatings on metals. *Prog Org Coat* 21:205–226
- Randles JEB (1947) Kinetics of rapid electrode reactions. *Discuss Faraday Soc* 1:11–19
- Raps D, Hack T, Wehr J et al (2009) Electrochemical study of inhibitor-containing organic-inorganic hybrid coatings on AA2024. *Corros Sci* 51:1012–1021
- Sagoe-Crentsil KK, Glasser FP, Irvine JTS (1992) Electrochemical characteristics of reinforced concrete corrosion as determined by impedance spectroscopy. *Br Corros J* 27:113–118
- Scuderi CA, Mason TO, Jennings HM (1991) Impedance spectra of hydrating cement pastes. *J Mater Sci* 26:349–353
- Scully JR (1989) Electrochemical impedance of organic-coated steel: correlation of impedance parameters with long-term coating deterioration. *J Electrochem Soc* 136:979–990
- Scully JR, Hensley ST (1994) Lifetime prediction for organic coatings on steel and a magnesium alloy using electrochemical impedance methods. *Corrosion* 50:705–716
- Sherif EM, Park S-M (2005) Effects of 1,5-naphthalenediol on aluminum corrosion as a corrosion inhibitor in 0.50 M NaCl. *J Electrochem Soc* 152:B205–B211
- Sherif EM, Park SM (2006) Effects of 1,4-naphthoquinone on aluminum corrosion in 0.50 M sodium chloride solutions. *Electrochim Acta* 51:1313–1321
- Shi Y, Tada E, Nishikata A (2015) A method for determining the corrosion rate of a metal under a thin electrolyte film. *J Electrochem Soc* 162:C135–C139
- Shkirskiy V, Volovitch P, Vivier V (2017) Development of quantitative local electrochemical impedance mapping: an efficient tool for the evaluation of delamination kinetics. *Electrochim Acta* 235:442–452
- Shreeram DD, Li S, Bedekar V et al (2017) Effect of reverse pulse time on electrodeposited Ni-W coatings. *Surf Coat Technol* 325:386–396
- Sığırçık G, Yıldırım D, Tüken T (2017) Synthesis and inhibitory effect of N,N'-bis(1-phenylethanol) ethylenediamine against steel corrosion in HCl media. *Corros Sci* 120:184–193
- Singh SK, Tambe SP, Gunasekaran G et al (2009) Electrochemical impedance study of thermally sprayable polyethylene coatings. *Corros Sci* 51:595–601
- Stafford OA, Hinderliter BR, Croll SG (2006) Electrochemical impedance spectroscopy response of water uptake in organic coatings by finite element methods. *Electrochim Acta* 52:1339–1348
- Su Q, Allahar KN, Bierwagen GP (2008) Application of embedded sensors in the thermal cycling of organic coatings. *Corros Sci* 50:2381–2389
- Su Q, Allahar KN, Bierwagen GP (2010) In situ embedded sensor monitoring of a United States Air Force primer beneath a topcoat exposed to atmospheric humidity and thermal conditions. *Corrosion* 66:066001-1–066001-12

- Takeuchi K, Nishijima A, Ui K et al (2005) Corrosion behavior of Fe-Cr alloys in  $\text{Li}_2\text{CO}_3\text{-K}_2\text{CO}_3$  molten carbonate. *J Electrochem Soc* 152:B364–B368
- Tan YJ, Bailey S, Kinsella B (1996) An investigation of the formation and destruction of corrosion inhibitor films using electrochemical impedance spectroscopy (EIS). *Corros Sci* 38:1545–1561
- Thee C, Hao L, Dong J et al (2014) Atmospheric corrosion monitoring of a weathering steel under an electrolyte film in cyclic wet-dry condition. *Corros Sci* 78:130–137
- van der Weijde DH, van Westing EPM, de Wit JHW (1994) Electrochemical techniques for delamination studies. *Corros Sci* 36:643–652
- van Westing EPM, Ferrari GM, de Wit JHW (1994) The determination of coating performance with impedance measurements-II. Water uptake of coatings. *Corros Sci* 36:957–977
- Vedalakshmi R, Manoharan S, Song HW, Palaniswamy N (2009) Application of harmonic analysis in measuring the corrosion rate of rebar in concrete. *Corros Sci* 51:2777–2789
- Vera Cruz RP, Nishikata A, Tsuru T (1996) AC impedance monitoring of pitting corrosion of stainless steel under a wet-dry cyclic condition in chloride-containing environment. *Corros Sci* 38:1397–1406
- Walter GW (1991a) The application of impedance spectroscopy to study the uptake of sodium chloride solution in painted metals. *Corros Sci* 32:1041–1058
- Walter GW (1991b) The application of impedance methods to study the effects of water uptake and chloride ion concentration on the degradation of paint films-I. Attached films. *Corros Sci* 32:1059–1084
- Wittmann MW, Leggat RB, Taylor SR (1999) The detection and mapping of defects in organic coatings using local electrochemical impedance methods. *J Electrochem Soc* 146:4071–4075
- Wu YM, Rapp RA (1991) Electrochemical impedance studies of hot corrosion of preoxidized Ni by a thin-fused  $\text{Na}_2\text{SO}_4$  film. *J Electrochem Soc* 138:2683–2690
- Yadav AP, Nishikata A, Tsuru T (2004a) Electrochemical impedance study on galvanized steel corrosion under cyclic wet-dry conditions-influence of time of wetness. *Corros Sci* 46:169–181
- Yadav AP, Suzuki F, Nishikata A, Tsuru T (2004b) Investigation of atmospheric corrosion of Zn using ac impedance and differential pressure meter. *Electrochim Acta* 49:2725–2729
- Yang BY, Kim KY (1999) The oxidation behavior of Ni-50%Co alloy electrode in molten Li+K carbonate eutectic. *Electrochim Acta* 44:2227–2234
- Yeh TK, Chien YC, Wang BY, Tsai CH (2008) Electrochemical characteristics of zirconium oxide treated type 304 stainless steels of different surface oxide structures in high temperature water. *Corros Sci* 50:2327–2337
- Zebarast HR, Rogak S, Asselin E (2013) Electrochemical detection of corrosion product fouling in high temperature and high pressure solution. *Electrochim Acta* 100:101–109
- Zeng CL, Li J (2005) Electrochemical impedance studies of molten  $(0.9\text{Na}, 0.1\text{K})_2\text{SO}_4$ -induced hot corrosion of the Ni-based superalloy M38G at 900 °C in air. *Electrochim Acta* 50:5533–5538
- Zeng CL, Wang W, Wu WT (2000) Electrochemical-impedance study of the corrosion of Ni and FeAl intermetallic alloy in molten  $(0.62\text{Li}, 0.38\text{K})_2\text{CO}_3$  at 650 °C. *Oxid Met* 53:289–302
- Zeng CL, Wang W, Wu WT (2001) Electrochemical impedance models for molten salt corrosion. *Corros Sci* 43:787–801
- Zeng C, Guo P, Wu W (2004a) Electrochemical impedance of two-phase Ni-Ti alloys during corrosion in eutectic  $(0.62\text{Li}, 0.38\text{K})_2\text{CO}_3$  at 650 °C. *Electrochim Acta* 49:2271–2277
- Zeng CL, Guo PY, Wu WT (2004b) Electrochemical impedance spectra for the corrosion of two-phase Cu-15Al alloy in eutectic  $(\text{Li}, \text{K})_2\text{CO}_3$  at 650 °C in air. *Electrochim Acta* 49:1445–1450
- Zeng Z, Lillard S, Cong H (2016) Effect of salt concentration on the corrosion behavior of carbon steel in  $\text{CO}_2$  environment. *Corrosion* 72:805–823
- Zhang JT, Hu JM, Zhang JQ, Cao CN (2004) Studies of impedance models and water transport behaviors of polypropylene coated metals in NaCl solution. *Prog Org Coat* 49:293–301
- Zhao Y, Zhou E, Liu Y et al (2017) Comparison of different electrochemical techniques for continuous monitoring of the microbiologically influenced corrosion of 2205 duplex stainless steel by marine *Pseudomonas aeruginosa* biofilm. *Corros Sci* 126:142–151

- Zheludkevich ML, Yasakau KA, Poznyak SK, Ferreira MGS (2005) Triazole and thiazole derivatives as corrosion inhibitors for AA2024 aluminium alloy. *Corros Sci* 47:3368–3383
- Zhou Q, Wang Y (2013) Comparisons of clear coating degradation in NaCl solution and pure water. *Prog Org Coat* 76:1674–1682
- Zhou Q, Wang Y, Bierwagen GP (2012) Influence of the composition of working fluids on flow-accelerated organic coating degradation: deionized water versus electrolyte solution. *Corros Sci* 55:97–106
- Zhou Q, Wang Y, Bierwagen GP (2014) Flow accelerated degradation of organic clear coat: the effect of fluid shear. *Electrochim Acta* 142:25–33
- Zhu B, Lindbergh G, Simonsson D (1999) Comparison of electrochemical and surface characterisation methods for investigation of corrosion of bipolar plate materials in molten carbonate fuel cell: part I. Electrochemical study. *Corros Sci* 41:1497–1513
- Zhu C, Xie R, Xue J, Song L (2011) Studies of the impedance models and water transport behaviors of cathodically polarized coating. *Electrochim Acta* 56:5828–5835
- Zhu J, Xu L, Lu M (2015) Electrochemical impedance spectroscopy study of the corrosion of 3Cr pipeline steel in simulated CO<sub>2</sub>-saturated oilfield formation waters. *Corrosion* 71:854–864

---

## **Part IV**

### **Electromagnetic Microwave and Millimeter-Wave Techniques**



# Microwave and Millimeter Wave Nondestructive Testing and Evaluation

26

Nathan Ida

## Contents

Historical Perspective .....	930
Overview .....	931
Energy and Safety Associated with Microwaves .....	934
Basics .....	935
Some Theory .....	935
Instrumentation .....	938
Interaction with Materials and General Setups .....	939
Microwave Microscopy .....	945
Resonant Testing Methods .....	950
Testing in Microwave Cavities .....	953
Monitoring of Moisture .....	953
Other Methods .....	957
UWB Radar .....	957
Microwave Radiometry .....	959
References .....	960

### Abstract

Microwave and millimeter wave NDE is an established method of evaluation with considerable advantages in areas that it applies. While mostly limited to lossy and lossless dielectrics, it has also found applications in surface analysis on conducting media. The number of applications is vast both in established areas of NDE and in emerging methods. The work and methods reported here are limited to the newer emerging methods but also to those that offer unique advantages such as microwave microscopy and in particular resonant methods where the emphasis is on open resonators and industrial scale testing and evaluation. Some common uses of microwaves for testing such as ground-penetrating radar and the

---

N. Ida (✉)

Department of Electrical and Computer Engineering, The University of Akron, Akron, OH, USA  
e-mail: [ida@uakron.edu](mailto:idak@uakron.edu)

use of microwaves in thermography are omitted since these are discussed elsewhere. A discussion of the emerging methods of ultra-wideband radar (UWB) testing and radiometry concludes the chapter.

---

## Historical Perspective

From a historical point of view, microwave NDE can be traced back to about 1945 with its first known application to inspection of dielectric materials for their moisture content and foreign objects in sheet dielectrics (Liskow 1948). Although microwaves may be considered a latecomer in comparison with other methods, their use in NDE occurred only a few years after the first microwave systems and equipment came into being following the invention of the cavity magnetron in 1939 and klystron in 1937 (Ginzton 1975) (various types of microwave devices can be traced back to the early 1920s, but the klystron and the magnetron were the first that were produced commercially). Even the waveguide did not exist before 1932. However, the theory of microwaves goes back to at least 1887 (Lord Rayleigh 1897). The same year, Heinrich Hertz, in his experimental proof of propagation of electromagnetic waves, produced waves at 0.6 m (about 500 MHz), well within the microwave frequency range (Cassidy et al. 2002) (Hertz used a pulsed source, and although the fundamental frequency of the pulses is not known, it is estimated anywhere between 50 MHz and 1 GHz). As equipment became more widely available and the understanding of the interaction of microwaves with materials increased, so did applications of microwaves for nondestructive testing and evaluation. It is common to mark the beginning of this trend with a paper by Hochschild (Hochschild 1963), but it should be remembered that the progression was gradual with many applications and publications before 1963 (Leonard and Stropki 1961; Soohoo 1962).

Initial applications of microwave techniques to NDE came slowly and tentatively by taking advantage of the effects substances such as water and variations in properties such as density or thickness have on propagating waves. These attempts were a direct outcome from measurements and work on radar. Only later, starting in the early 1970s, was it recognized that microwaves could be used to detect flaws such as voids and delaminations in composites (Decreton and Gardiol 1974; Bachtieri et al. 1994; Ghodgaonkar and Ali 2000) and for testing of ceramics (Bahr 1978). With the increased acceptance of microwave testing came a new generation of researchers that advanced the science behind and the application of microwave NDE. The first microwave devices can be traced to the use of the crystal diode in the early 1930s, but the real impetus came in the 1960s with the development of negative resistance devices and field-effect transistors (*FETs*). The use of gallium arsenide (*GaAs*) and indium phosphide (*InP*) have also benefited this aspect of solid-state device development by advancing the frequencies at which these devices could operate and eventually lead to microwave integrated circuits.

Passive devices were developed in parallel with the development of microwave sources. These include a variety of antennas, couplers, filters, attenuators, waveguides, cavity resonators, absorbers, terminators, rotators, and many others.

Microwave circuit development followed steps similar to other circuits at lower frequencies. As examples, the idea of heterodyne receivers was used starting with the earliest radar equipment, while microwave integrated circuits find applications in many communication systems.

Developments in microwave equipment and in particular the availability of the network analyzer starting in the 1970s have also contributed greatly to this trend so that by the late 1970s the method was recognized as a viable and useful method for a wide array of tests and measurements in a variety of materials with emphasis on dielectrics (Hochschild 1963; Ash and Nichols 1972; Bachtiari et al. 1994).

A common thread in the development of microwaves for NDE is that much of that happened in conjunction with communication and the NDE/NDT community has gained considerably from this synergetic link. It meant that equipment as well as processing algorithms were largely available and could be adapted to the NDE environment. That is not to say that the adaptation is trivial. For example, ground-penetrating radar, a common NDT/NDE method for assessment of concrete, borrowed freely from radar principles and processing methods, but the method today resembles very little a classical radar system (Joongsuk and Nguyen 2005; Han and Nguyen 2007; Yang et al. 2017; Travassos 2007).

---

## Overview

Microwave nondestructive evaluation is the art of testing of materials and quantifying the results of testing using microwaves as the source of energy. Unlike many other NDE methods, which were developed specifically for testing, microwave NDE is an extension and appropriation of techniques that have matured in disciplines as diverse as radar and communication but also in radiometry, antennas, and methods of analysis, processing, and imaging. As such the method's development is closely linked with developments in diverse areas of microwave work and with components, instrumentation, and methods associated with that work. That also means that often, a method of microwave NDE is simply an application of a measurement, sensing, or detection method for the purpose of testing and evaluation.

Testing with microwaves is dominated by the basic properties of microwaves. Since their penetration in conducting materials is minimal, they are mainly used to test nonconducting materials. This includes dielectric and lossy dielectric materials. Testing and measurements on conducting materials are limited to dimensional testing such as thickness gaging and surface measurements such as testing for surface-breaking flaws, effects on the surface such as corrosion, and surface roughness and surface flaws including those undercoatings. But microwaves are affected by a large number of material properties related to the basic electrical properties: conductivity, permittivity, and permeability. In lossless or lossy dielectrics, porosity, material composition, uniformity of the material, delamination of layers, moisture, and contamination content are only some of the properties that can be measured. The range of nonmetallic materials in which this is possible is extensive and growing steadily. This includes ceramics, plastics, polymers and composites, concretes, as

well as organic materials such as wood products, foods, or biological materials. The spatial resolution that can be expected of microwave tests depends on the wavelength of the wave. For microwaves and millimeter waves, the wavelength ranges between 1 m for the lowest microwave frequency (300 MHz) and 1 mm for the highest microwave frequency (300 GHz). This resolution indicates the ability of the test to discern closely spaced discontinuities in the materials at its most basic level. However, more sensitive measurements are possible by correlating them with changes in amplitude or phase and through a variety of techniques that include microwave microscopy, evanescent waveguide methods, and a variety of resonant techniques. This is the case with dimensional measurements where variations in thickness of a few microns are measurable even at the lower end of the microwave spectrum. High-resolution testing can be achieved by scanning of the microwave field and by moving the source. This is similar to synthetic-aperture radar (SAR) methods and can be used for imaging. If this is done in the near field of the antenna, it may even be called "micrometry."

Another particular property of microwave testing is the means by which energy is coupled into the testing environment. This can be as simple as a horn antenna or an open microwave guide. In some other cases, an aperture is used for this purpose, or in the case of microwave cavities, a simple probe or loop serves the purpose. In all cases, the coupling can be done through air or a convenient dielectric, and typically no special treatment or coupling media are needed. While many tests are done in what may be termed the "near-field" environment (close to the antenna or source), measurements in the far field are also possible where the waves propagate through a medium such as air. This is particularly applicable to scattering methods, including radar testing applications.

Because of the influence of so many effects and properties on the wave, the testing environment can be, and often is, noisy. Reflections from near and far surfaces, edges, and other artifacts in the material are often encountered, resulting in noise and loss of resolution. Microwave radiation is highly directive, and, because of the short wavelengths involved, the devices used are often very compact. While many of the applications are in high power communication and radar system, low power applications are just as common.

Of primary interest in this work is the interaction of microwaves with materials and components. This takes the form of attenuation (scattering and absorption) in materials, reflection, and resonance. These effects are exploited in various testing arrangements to allow for quantitative measurements in materials. The absorption of microwaves in water is well known and widely used in microwave ovens. These ovens depend on this effect because it is typical of the whole microwave range. Thus, while most microwave ovens operate at specific frequency bands because of regulation, they can also operate at other frequencies. The absorption in water can be used either directly or indirectly for testing of moisture and related effects (e.g., curing, drying, or amount of substance). Other materials absorb radiation at specific frequencies. Often, these are narrowbands that allow chemical analysis of materials. These resonant frequencies are extremely useful for material characterization and identification.

Sometimes even traces of materials can alter the resonant frequency of a microwave cavity, leading to the detection of materials in trace amounts. This again may be used for testing or detection. Typical applications of this type are contamination tests and detection of explosives. Because microwaves border on the one hand the high-frequency radio range and on the other the low infrared range, they have properties of both. More than any other frequency range, the microwave range is sometimes analyzed using circuit theory and sometimes using wave theory. Transmission lines are almost always analyzed as distributed parameter circuits, while the aspects of refraction, transmission, and propagation in waveguides are analyzed using wave theory.

The methods that can be brought to bear for NDE in the microwave range are truly vast. The most obvious are those related to propagation of electromagnetic waves: reflection, transmission, and scattering of waves (Kharkovsky et al. 2002). These are often used for evaluation of material properties (Stuchly and Stuchly 1980; Arunachalam et al. 2006; Hughes and Zoughi 2005a, b; Mukherjee et al. 2016; Bois et al. 2000; Peer et al. 2003; Mukherjee et al. 2018) but also for dimensional measurements and detection (Caorsi et al. 2003; Bogosanovic et al. 2012). Scattering and absorption of electromagnetic waves can also be included in this group. Then there are various resonant methods that can and are being used. But there are others. Radar-like methods, including Doppler and SAR techniques, are also applicable. A particularly interesting method is the ground-penetrating radar (GPR), developed specifically for microwave assessment of dielectrics and lossy dielectrics such as soils and concrete (Joongsuk and Nguyen 2005; Han and Nguyen 2007; Yang et al. 2017; Travassos 2007). The heating effects of electromagnetic radiation, loading of open-ended waveguides, or simple probes (antennas) can also be used for various tests (Bakhtiari et al. 1993; Mazlumi et al. 2006; Qaddoumi et al. 2000; Kharkovsky et al. 2011; Jundi and Qaddoumi 2009). Testing and evaluation can be done in the near field (Ghasr 2004; Qaddoumi et al. 2007; Akuthota et al. 2004; Fear et al. 2002a, b), leading to methods that can be identified as microwave microscopy with resolutions well below a fraction of a wavelength (Tabib-Azar et al. 2002; Rosner and Van der Weide 2002; Joffe et al. 2017; Ash and Nichols 1972; Anlage et al. 2007; Ciocan 2000; Ciocan and Ida 2004; Chen et al. 2005) as well as in the far field (Arunachalam et al. 2006; Mukherjee et al. 2018, MTHUU-1). Detection and evaluation can be done on the surface of materials or in the bulk with low or high penetration depending on material properties and frequency (Xu et al.). Imaging, tomography (Broquetas et al.), and inversion (Mukherjee et al. 2016, 2018; Lerossey et al. 2004; De Rosny et al. 2010; Reyes-Rodríguez et al. 2014; Liu et al. 2005) are also possible in dielectrics and composites, and resonant methods are often used for material evaluation and moisture content studies (Ida 2008, 2018; Ida and Bhuyia 2008; Auld 1978; Li et al. 2011; Joffe et al. 2017). In terms of applications, microwaves are best suited for evaluation in dielectrics and low-loss dielectrics such as ceramics, rubber, and plastics (Ganchev et al. 1994; Decretion and Gardiol 1974; Bachtiai and Zoughi 1990) but also in higher loss materials such as concrete, composites, sands, and carbon fiber-based products (Kharkovsky et al. 2002; Bois et al. 2000; Trabelsi et al. 1997; Yang et al. 2017) and in biological media (Stuchly and Stuchly 1980; Fear et al. 2002a, b; Winters et al. 2006; Mehta et al. 2006).

Applications in characterization of dielectric and magnetic properties can take various forms in dielectrics, magnetic materials, mixtures of solid or liquid forms, glasses, foams, resins, and others including concretes of various types. Properties can be monitored for a variety of conditions including moisture content (Ida 2008, 2018; Ida and Bhuyia 2008), porosity, and consistency (Ganchev et al. 1994; Bois et al. 2000; Hughes and Zoughi 2005a, b; Peer et al. 2003; Trabelsi et al. 1997; Mubarak et al. 2001; Decreton and Gardiol 1974). Microwave testing can be used for dimensional measurements (Bachtiari et al. 1994; Anderson 1997; Sayar and Ogawa 2009; Li et al. 2011; Ghasr et al. 2015; Zoughi et al. 2016) with resolutions down to micrometers (Bachtiari and Zoughi 1990; Ganchev et al. 1995; Qaddoumi et al. 2002), detection of conditions such as disbonds and delaminations (Bachtiari et al. 1994; Ganchev et al. 1995; Qaddoumi et al. 1996, 2002; Gray and Zoughi 1997), as well as flaws on surfaces of conductors (Mazlumi et al. 2006; Huber et al. 1997). Applications to metals also include detection of some fatigue flaws, surface roughness, and evaluations of coatings such as paint (Bahr 1981, 1982) as well as corrosion effects on surfaces or undercoatings (Qaddoumi et al. 1997, 2000; Hughes et al. 2001; Ghasr et al. 2005a, b; Mast 2001; Kharkovsky et al. 2006).

Although this would indicate a vast number of effects and applications, these are only a fraction of what is possible with microwave NDE. However in principle most tests can be summarized by the evaluation of permittivity, permeability, and conductivity and anything that affects these three properties, including dimensional parameters.

---

## Energy and Safety Associated with Microwaves

The radiation energy associated with microwaves can be estimated considering the quantum equivalent photon. The energy of a photon is equal to  $hf$  where  $h$  is the Planck constant ( $h = 4.14 \times 10^{-15}$  ev). Thus the maximum energy of a photon in the microwave range is roughly  $1.2 \times 10^{-3}$  ev (minimum is about  $1.0 \times 10^{-6}$  ev at the lower frequency range). This energy is relatively low and is much lower than the energy needed for ionization, that is, the energy is much lower than the energy in molecular links. Thus, because it cannot break these links, it is considered a nonionizing form of radiation.

The danger, if any, from microwave radiation is considered to be primarily due to absorption (causing heating) rather than due to its intrinsic energy. For this reason, the safety levels of radiation are defined on the surface, in terms of power per unit area ( $\text{W/m}^2$ ) or ( $\text{mW/cm}^2$ ), in terms of specific absorption rates (SAR) in  $\text{W/kg}$ , or in terms of electric field intensity [ $\text{V/m}$ ]. Limits on exposure vary from country to country, but in general they follow standards set by various bodies including the World Health Organization (WHO) (WHO 2016), the International Commission on Non-Ionizing Radiation Protection (ICNIRP) (ICNIRP 1998), the International Electrotechnical Commission (IEC) (ITU K.52 2016), the European Committee for Electrotechnical Standardization (CENELEC), national regulatory bodies such as the Federal Communications Commission (FCC) in the US (OET Bulletin 65 1997) standards organizations (IEEE 2002, 2006), and other interested bodies

(International Commission on Non-Ionizing Radiation Protection 2016; ACGIH 2017; SCENIHR 2015). There is some evidence that nonthermal effects of microwave radiation also play a role in safety, but this issue is still controversial and not very well researched. The exposure levels allowed should serve as guidelines only. While there may be no harmful effects due to allowable levels of microwave radiation, as a matter of prudence, one should avoid all exposure to microwave radiation if only because of its absorption effects.

## Basics

### Some Theory

Microwave testing and evaluation is based on the properties of electromagnetic waves and their interaction with materials in the microwave frequency range. That covers the whole range of effects and interactions from simple propagation in lossless or lossy materials, reflection and transmission of waves across material boundaries, attenuation and phase changes, antennas, and many more including propagation in waveguides and transmission lines and resonance in microwave cavities or transmission line resonators. It is therefore impractical to discuss here the whole electromagnetic theory (Ida 2015). There are however a few concepts that are fundamental to testing, and these are summarized here, assuming the reader is either familiar with the concepts or will undertake their study as necessary. The main reason to discuss these here is to list the concepts that are being used in this chapter.

For a wave to exist and be able to carry power, it must exhibit an electric field intensity  $\mathbf{E}$  [V/m] and a magnetic field intensity  $\mathbf{H}$  [A/m]. Both  $\mathbf{E}$  and  $\mathbf{H}$  are vector quantities, and for a wave to carry power, their vector product must be nonzero. The time-averaged power density (real power density) in the wave is:

$$\mathbf{P}_{av} = \frac{1}{2} \operatorname{Re}\{\mathbf{E} \times \mathbf{H}^*\} \quad \left[ \frac{\text{W}}{\text{m}^2} \right] \quad (1)$$

Power density is a vector as well and indicates both the magnitude and direction of propagation of power. An equivalent relation for propagation of power on a transmission line replaces  $\mathbf{E}$  and  $\mathbf{H}$  by  $V$  and  $I$  and produces power on the line:

$$P_{av} = \frac{1}{2} \operatorname{Re}\{VI^*\} \quad [\text{W}] \quad (2)$$

The ratio between the magnitudes of  $\mathbf{E}$  and  $\mathbf{H}$  is called the wave impedance or the intrinsic impedance in which the wave propagates. For the common TEM modes of propagation (see below), the wave impedance is:

$$\eta = \frac{|\mathbf{E}|}{|\mathbf{H}|} \quad [\Omega] \quad (3)$$

The characteristic impedance of transmission lines is defined using similar concepts and normally denoted as  $Z_0$ .

Waves propagate subject to a propagation constant  $\gamma = \alpha + j\beta$  where  $\alpha$  is the attenuation constant and  $\beta$  is the phase constant. Thus a wave propagating a distance  $d$  from the source will experience attenuation and a change in phase as follows (for TEM waves, see below):

$$E = E_0 e^{-\alpha d} \cos(\omega t - \beta d) \quad \text{or :} \quad E = E_0 e^{-\alpha d} e^{-j\beta d} \quad (4)$$

The first form is in the time domain; the second is written with phasors in the frequency domain. Identical forms apply to the magnetic field, and, in general,  $\mathbf{E}$  can be a function of space. Attenuation is exponential whereas the phase is linear. Both depend on conductivity, permeability, permittivity, and frequency. The quantity  $\beta d$  is called the electric length.

Waves may be transverse electromagnetic (TEM, both  $\mathbf{E}$  and  $\mathbf{H}$  are perpendicular to each other and to the direction of propagation), transverse electric (TE,  $\mathbf{E}$  is perpendicular to the direction of propagation, but  $\mathbf{H}$  has components perpendicular and in the direction of propagation), or transverse magnetic (TM,  $\mathbf{H}$  is perpendicular to the direction of propagation, but  $\mathbf{E}$  has components perpendicular and in the direction of propagation). The type of wave has consequences on all properties of the wave. For example, the wave impedance of TM waves is lower than that of TEM waves, whereas the wave impedance for TE waves is higher than the TEM wave impedance. In general, waves propagating in unbounded space may be considered as TEM waves, whereas waves propagating in waveguides are either TE or TM waves.

The wave propagates at a finite speed called phase velocity, which depends on material properties and mode of propagation. In its simplest form, in a perfect dielectric, the phase velocity may be written as:

$$v_p = \frac{1}{\sqrt{\mu\epsilon}} \quad \left[ \frac{\text{m}}{\text{s}} \right] \quad (5)$$

where  $\mu$  is the permeability and  $\epsilon$  is the dielectric constant of the medium in which the wave propagates. In most media, the dielectric constant is the real part of the permittivity of the material  $\epsilon_c$ :

$$\epsilon_c = \epsilon' + j\epsilon'' = \epsilon \left( 1 - j \frac{\sigma}{\omega\epsilon} \right) \quad (6)$$

where  $\sigma$  is the conductivity of the medium and  $\omega = 2\pi f$  where  $f$  is the frequency. The ratio between the imaginary and real part of permittivity is a measure of losses in the medium and is called the loss tangent:

$$\tan \theta = \frac{\sigma}{\omega\epsilon} \quad (7)$$

In real materials, the dielectric constant in (5) is replaced with  $\epsilon_c$ , and phase velocity depends on conductivity and becomes frequency dependent. It also depends

on the type of wave supported in the medium of interest. Permeability is also in general complex.

The wavelength  $\lambda$  associated with the wave is often a convenient quantity to invoke in analysis. It is measured in meters and may be written simply as:

$$\lambda = \frac{v_p}{f} \quad [\text{m}] \quad (8)$$

At the interface between two media with different material properties, part of the wave reflects with a reflection coefficient  $\Gamma$ , and part of it transmits across the interface with transmission coefficient  $T$ . Given incident electric field intensity  $E_{\text{inc}}$  propagating in medium 1,  $\Gamma$  and  $T$  are:

$$\Gamma_{12} = \frac{E_{\text{ref.}}}{E_{\text{inc}}} \quad T_{12} = \frac{E_{\text{trans.}}}{E_{\text{inc}}} \quad (9)$$

The reflection and transmission coefficients depend on the material properties of the two media, on the angle of incidence of the wave and on the polarization of the wave (direction of the electric field intensity) relative to the interface. In transmission lines, the reflection and transmission coefficients depend on the load and characteristic impedances, which in turn depend on material properties.

Propagation of waves in waveguides (hollow metallic structures of a defined cross-section) is governed by the basic relations, with the additional effects of conducting surfaces including reflections of these surfaces, attenuation due to losses in the walls, and geometry of the waveguide. The waves in waveguides are either TE or TM modes propagating with a phase and attenuation constant that again are different than propagation in unbounded space. Unlike propagation in unbounded space or on transmission line, TE or TM modes can only propagate above a certain frequency, called the cutoff frequency, which depends on the mode and structure of the waveguide. Below that frequency the waves are highly attenuated to the point that propagation is not possible.

On most transmission lines, propagation is in TEM mode and can occur at all frequencies. Propagation properties on transmission line are defined similarly to those in space but are typically related to circuit parameters (resistance, capacitance, inductance, and conductance), all defined per unit length of the transmission line and dependent on dimensions and structure of the line and material properties of the line.

Waveguides and transmission lines can also oscillate by simply enclosing a waveguide to form a cavity or by shorting a transmission line. The enclosed structure causes reflections of the propagating waves, and these will interfere constructively to resonate at the frequencies at which such interference occurs. The resonant frequency depends on the dimensions and structure of the resonator and, most critically, on the properties of materials within the resonator.

One of the more common elements of most microwave system and, indeed, of many electromagnetic systems is the antenna. Its purpose is to generate a wave that can then propagate in a medium surrounding the antenna. In fact the concept is very simple – a conductor of a defined shape carrying a time-dependent current will

generate around itself an electric and a magnetic field and hence, based on the relations above, a wave. The antenna can take the form of a cylindrical conductor or a loop. Of course, not all antennae are conductors carrying currents. An antenna can be an opening (an aperture) in a waveguide or a cavity resonator (an aperture antenna) or may be a flared termination of a waveguide (a horn antenna). What is important is that the antenna can generate an electric and magnetic field such that the vector product of the two is nonzero. The physical size of the antenna defines its frequency response. In general antennas are very efficient in transmitting power in spite of their simplicity.

In terms of operation, one distinguishes between the near field and the far field of antennas. In the near field, which is defined by distances smaller than a wavelength, antennas behave either as capacitive or inductive devices (depending on the type of antenna). As such, there is no radiation, but there is coupling of energy into the near field. In most cases therefore, the near field of antennas is of little interest except in applications where one needs to couple energy into a system. Examples of these are the coupling of energy into waveguides or cavity resonators. In such cases, a short wire or a loop serve the purpose although they are usually called probes rather than antennas. The far field of antennas is that region that is much further than a wavelength (at the frequency the antenna operates) and is characterized by radiation based on the relations above.

Antennas come into a bewildering variety of types, sizes, and shapes. They may be designed for specific applications or general purpose and may be directional or omnidirectional, low or high power, resonant or broadband, and so on. Antennas are often integral with circuits and may be no more than a trace on a printed circuit board or may be massive installations of reflectors and control machinery to track signals. However, they all operate on essentially the same principles. In addition, all antennas can serve as transmitting or receiving elements based on the duality principle in antenna theory (Stutzman and Thiele 2013; Balanis 2005; Ida 2018).

## Instrumentation

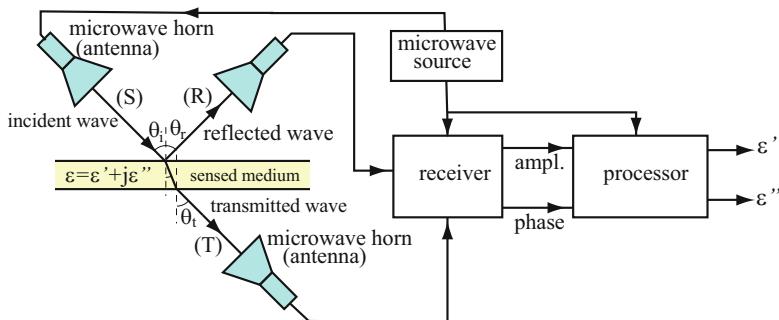
One of the perennial issues associated with microwave NDE is that of instrumentation and circuits. When one thinks of a circuit, the classical electric circuits come to mind. But at microwave frequencies, electronic circuits are less useful than at lower frequencies. Often one deals with devices that are foreign to most test technicians. Signals are generated using microwave tubes or through harmonic generators and are propagated over transmission lines, over waveguides, or through air. Coupling of power into a test environment is done with an antenna (classical antennas or probes, horn antennas, apertures in waveguides, or cavity-backed apertures). Detection and quantification of power may be done with diode detectors, thermistors, bolometers, or, in some cases, using calorimetric methods, especially at higher power levels. Simple components such as attenuators, phase shifters, and couplers are likely to be made of various types of transmission line circuits. Then there are circuits and components unique to microwaves such as directional couplers, splitters, isolators,

and circulators. All of these are available commercially and allow the practitioner to assemble the circuits required for useful tests. As a consequence, the resulting instruments tend to be bulky with little flexibility in construction and often quite expensive. Mechanical assembly of the various components is required and must be exact to eliminate losses and unwanted signals due to reflections of waves from discontinuities, poor connections, and cables. Mastering microwave devices is a rather specialized endeavor that takes time and resources and must be undertaken very carefully.

Fortunately one can lighten some of these issues through the use of network analyzers. A network analyzer is a specialized instrument of considerable complexity, but in its most fundamental form, it can be thought of as a microwave source that can be coupled to the test environment by some means, a detection mechanism that can detect and quantify signals received from the environment and a computer to drive the measurement and compute measurement parameters. In practice it is much more than that. The source itself can be scanned over a large frequency range and at exact and variable power levels with great accuracy. The measurement can be any of a number of different possibilities, some of which are pre-programmed into the onboard computer, but in most cases, it is based on the so-called S-parameters. The latter are the reflections from both ports of the network ( $S_{11}$  and  $S_{22}$ ), transmission from the input to the output ( $S_{21}$ ) and from the output to the input ( $S_{12}$ ) (Rohde and Schwarz 2007; Wu 2007). These four parameters allow complete analysis of a linear two-port network connected to the analyzer. From these, one can obtain, again with the aid of the analyzer, other parameters and test results including resonant frequencies (Ida 2018; Das and Das 2000; Pozar 1998), frequency response, material properties (Weir 1974; Nicholson and Ross 1970; Kadaba 1984), and many more (Dunsmore 2012; Godgaonkar et al. 1990). Vector analyzers are either scalar (measure amplitudes only) or vector (measure amplitudes and phase) and in frequency ranges from sub-GHz to well in excess of 150 GHz. Although the typical measurement assumes a linear network, nonlinear measurements are also possible as are measurements in space, where the network is not clearly defined.

## Interaction with Materials and General Setups

As mentioned in the introduction, some of the most common and most useful microwave methods rely on the reflection, transmission, or scattering of microwaves. These include simple arrangements of sources, antennas, and detectors or may be as complex as a radar system. In all cases, however, advantage is taken of the fact that the reflection, transmission, or scattering of electromagnetic waves is affected by the properties of the medium and hence may be used to evaluate its state including evaluation of material properties and conditions of the medium. The basic methods of reflection and transmission are shown in Fig. 1 in which both methods are used to analyze the complex permittivity of a medium. The incident wave is transmitted at an angle, and both the reflected and transmitted waves are used to evaluate the dielectric constant and the loss tangent of the medium through the use



**Fig. 1** A generic method of evaluation of the complex permittivity of a medium

of amplitude and phase measurements. The instrumentation shown can be and often is replaced by a network analyzer for simplified and accurate measurements. In many tests, only the reflected or only the transmitted waves are required, and these can be accomplished through a variety of arrangements as shown schematically in Figs. 3 and 7.

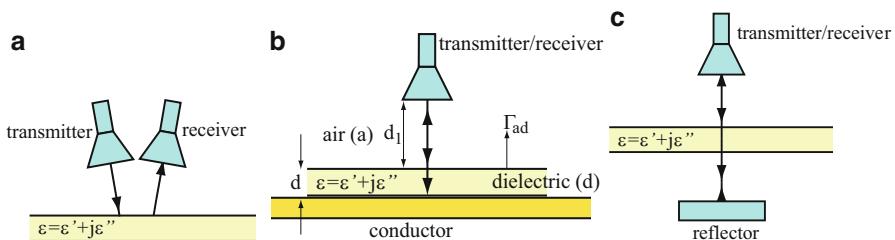
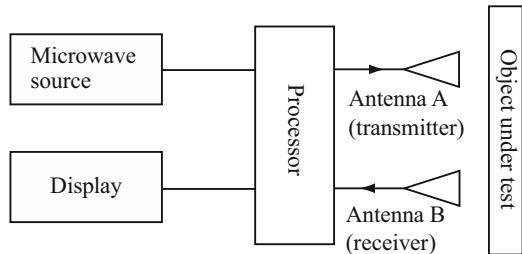
### Reflection Tests

A basic reflection test is shown in Fig. 2. The incident wave is produced by a source and coupled through an antenna. The waves are reflected off the material either completely (as for conductors) or partially (as for dielectrics and lossy dielectrics). The reflected wave is received by an antenna (antenna *B* in this case). The processing module in Fig. 2 identifies the two parts of the wave and produces the required display. This may be any of the effects on which reflection depends. In most cases, the changes in the dielectric constant or changes in losses (loss tangent) are correlated to a particular property. This may be as simple as thickness of the material, variations in dielectric constant (e.g., delamination), moisture content, contamination, and a variety of other parameters. In fact, almost any material property can be identified with some change in the reflection coefficient. In many cases, the source processor and display are accomplished through the use of a network analyzer, in the present application operating in reflection mode.

For simplicity, we have used two antennas: one for transmission and the other for reception. However, any transmitting antenna can also be used for reception, and in pulsed mode, the same antenna can serve both functions. The separation of the two waves can be done by directional couplers or, in some cases, not done at all, and measurements are performed on the total wave. There are many variations on this simple test, but in all cases one can distinguish between an incident and a reflected wave.

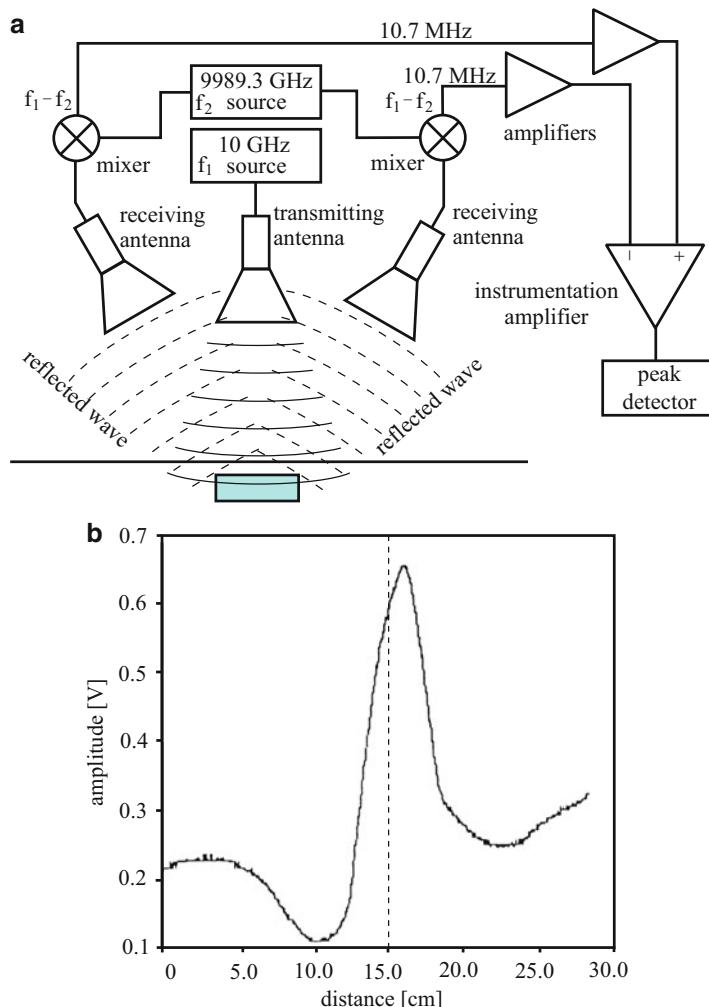
Three practical methods of reflection tests on dielectrics are shown in Fig. 3. The configuration in Fig. 3a or a variation of this can also be used on conducting surfaces. Figure 3b shows the use of a reflection test for coating and undercoating effects.

A simple example of the use of reflected waves for the detection of dielectric inclusions in a lossy medium is shown in Fig. 4a. This particular application relates

**Fig. 2** A basic reflection test**Fig. 3** Reflection tests on a dielectric. (a) Both antennas on one side, (b) reflection test on a conductor-backed dielectric, (c) the use of a reflector to reflect the signal

to detection of nonmetallic, antipersonnel mines buried in shallow soil. The central antenna transmits a 100 mW microwave at 10 GHz, from a height of 35 cm above the surface of the ground. The transmitted signal is partially reflected by the ground surface and partially transmitted into the ground and again partially reflected by the dielectric. The reflections are received by two antennas feeding two receivers after heterodyning. In this simple application, the output is simply the difference between the outputs of the two channels. The measurement in this case is greatly simplified by the fact that the primary function is detection rather than evaluation or imaging and by the fact that the measurement is differential and hence common mode effects such as ground clutter and noise are practically eliminated. The signal in Fig. 4 shows that the dielectric is approximately at the center of the 30 cm horizontal scan. The wavelength at 10 GHz is 3 cm, and one would expect to be able to detect the edges of the dielectric provided it is larger than a wavelength. That in fact is the case if the antennas are in much closer proximity (say at ground level). In this case, the need to clear obstacles on the ground dictated a higher clearance, and the edges can be seen as the broad dips on the two sides of the peak. The diameter of the dielectric is 12 cm corresponding well with the two dips even though they are not sharp.

In most testing applications, a more robust method is needed, that is, one requires exact values and correlation of these values with a property or condition in the material. In such cases, a network analyzer is likely to be used for both the transmitter and receiver, and the correlation is done through the measurement of the S-parameters. A schematic example is shown in Fig. 5, which shows a free-space measurement of permittivity using either a reflection test or, for that matter,

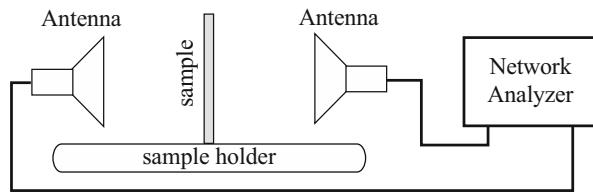
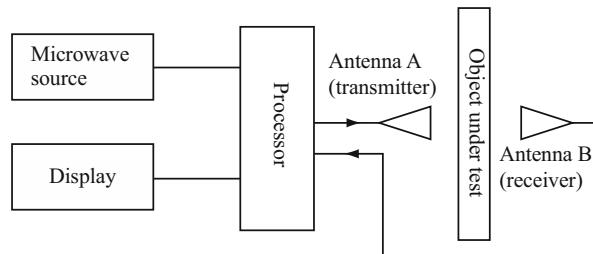


**Fig. 4** (a) Detection of buried objects using a differential reflection method. (b) The signal obtained shows a peak at the center of the dielectric and two dips indicating its corners

a transmission test. The network analyzer measures all four S-parameters from which the reflection and transmission coefficients are calculated, and these correlated with (primarily) material properties. This approach can be used in many configurations, and because the network analyzer can remove undesirable effects such as those caused by the sample holder, measurements can be very accurate.

### Transmission Tests

Transmission tests rely on measurements on the transmitted wave. This is shown schematically in Fig. 6 (see also Figs. 1 and 5). The transmission coefficient also

**Fig. 5** Transmission test**Fig. 6** A basic transmission test

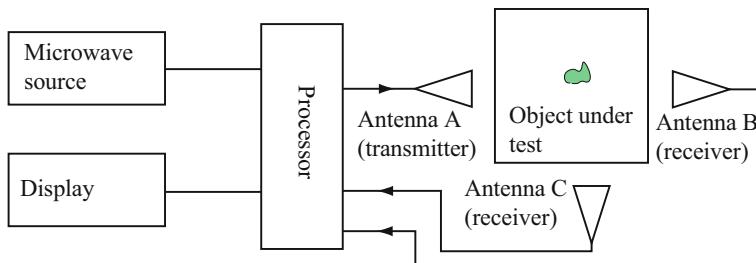
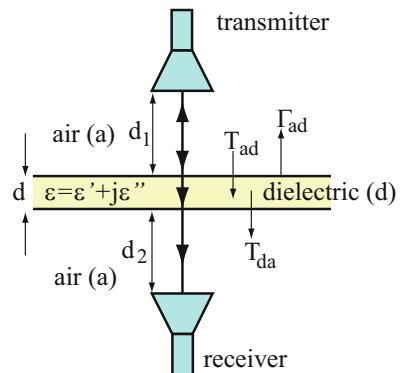
depends on material properties as the reflection coefficient. However, transmission tests are sometimes easier to perform or to correlate to material conditions. Attenuation measurements are also transmission tests and are often used for evaluation of lossy dielectrics. The basic modules of the test are the transmitting and receiving antennas and associated instrumentation. Processing is similar to that done for reflection tests. There are some cases for which only one of the tests is practical. For tests on conductors (such as thickness gaging), only reflection tests are possible. Since the transmission is negligible, it is not possible to use it for testing. Similarly, if the reflection coefficient is very small, especially for low-loss dielectrics, the transmission test might be more effective (Fig. 7).

An example of a transmission test to sense a variety of effects including thickness, water content, losses, and variations in permittivity caused, say, by density variations, delamination, or bulk inclusions is shown in Fig. 1 with only the source (S) and transmitted (T) antennas present. In this configuration, amplitude and phase of the received signal are compared with a reference signal from which the transmission coefficient can be calculated leading to information on the material under test. With proper calibration accurate measurements are possible although, as was the case with transmission tests, only bulk properties can be analyzed.

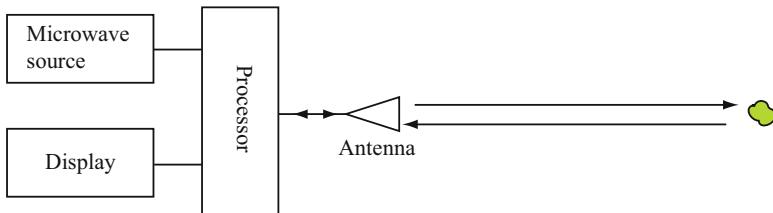
### Scattering Tests

Both reflection and transmission are properties associated with the bulk of the material. Scattering, as understood in the context of testing of material, on the other hand, is associated with local conditions in the material. While any variation in the material will affect both the reflection and transmission, we will view scattering as an indication of local effects such as flaws, inclusions, delaminations, surface roughness, etc. We also include radar and similar effects with scattering. The two basic tests associated with scattering are shown in Figs. 8 and 9. In Fig. 9, the scattered field is picked up by antenna C. This antenna will pick little or nothing

**Fig. 7** Transmission test on a dielectric



**Fig. 8** Scattering test for inclusions or flaws in dielectrics



**Fig. 9** Scattering test using a monostatic radar method

unless there is scattering within the test sample. Normally, only reflection and transmission will exist. Transmission is detected by antenna *B*, while reflection can be detected by antenna *A* or, perhaps, a separate antenna as in Fig. 1. This type of measurement can be used to detect scattering by foreign objects or sharp variations in material properties such as delaminations inclusions or interfaces between different materials. The second type is the basic radar system and is shown in Fig. 9. Here the pulsed radar method is shown although other radar methods can be used. In *NDT*, the pulses as scattered and picked up by the antenna are correlated with the location, size, or properties of the material.

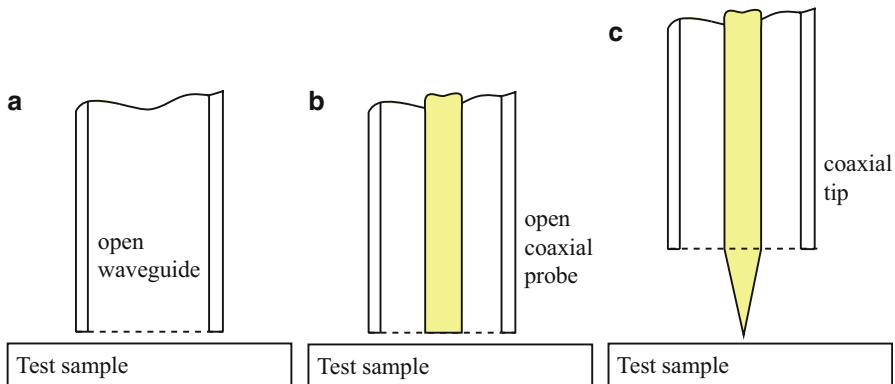
Scattering methods of particular interest are those associated with radar. Probably the best known of these methods is the ground-penetrating radar (GPR) which was mentioned in the introduction. This will not be discussed here, but it and its use in NDE is the subject of a separate chapter in this handbook. Nevertheless, it should be noted here that GPR is a common method of assessment of concrete structures whereby rebars, inclusions, cracks, delaminations, and other effects can be detected and imaged in near real time and over large areas. There are however uses of radar in NDE which, although similar or related to GPR, are nevertheless sufficiently different to be discussed here. One relatively new and useful radar in some applications is the ultra-wideband (UWB) radar, as discussed below.

---

## Microwave Microscopy

Microwave microscopy allows the quantitative measurement of properties below the scale of the wavelength expected from the wavelength. In most microwave systems, one can only detect variations in, say, material properties or dimensions of artifacts that are of the order of the wavelength, as was discussed in the introduction. However there are two methods that allow, under certain conditions, the detection and quantitative evaluation of properties well below the wavelength limit. One is to test the material in resonant cavities. In this approach, the shift in resonant frequency, especially when measured with high precision instruments such as network analyzers, can detect variations in dimensions of the order of micrometers at test wavelengths of the order of 1 m. This will be discussed in the following section. A second method, which is usually called microwave microscopy, involves the interaction of the microwave probing energy in the near field of the source where the spatial frequency is high (Anlage et al. 2007; Rosner and Van der Weide 2002; Joffe et al. 2017). In many cases this involves evanescent waves that interact with the material under test in a highly localized manner. There are a few additional conditions that must be satisfied. First, it is implicit that evaluation is done at relatively high frequencies. Second, because the very localized interaction, the test sample must be scanned in some pattern. Related to that, it is also clear that only small sections of a sample can be realistically scanned. Therefore the method is useful for localized testing and for that reason has been in the past limited to laboratory use and to specialized tests. Although the original development of microwave microscopy was to allow evaluation of material properties, especially permeability (Acher et al. 1996; Ustinov et al. 1999; Ciocan 2000), it can be used as well for dimensional tests such as surface flaws in conductors (Ash and Nicholls 1972; Tabib-Azar et al. 2002).

In most cases, microwave microscopy instruments consist of a sub-wavelength sensor or antenna that constitutes the microscopy in the instrument. This can something be as simple as an open waveguide, an open-ended transmission line, or a tip as shown in Fig. 10. In all three cases shown, use is being made of the fact that fields at an open waveguide or transmission line decay rapidly generating the required high spatial frequency. These features may be viewed as very small antennae. The fields from the aperture or tip interact with the sample. Some of the



**Fig. 10** Simple microwave microscopy probes. (a) Open waveguide. (b) Open coaxial probe. (c) Coaxial tip

energy in the fields is reflected; some is scattered. One monitors either the reflected fields or the scattered fields as a function of probe position to generate a measurement or an image of the sample properties. This type of probe is called a nonresonant probe. The analysis of the fields due to any of the probes in Fig. 10 can be done numerically using appropriate simulation software. Approximate analysis (Anlage et al. 2007) relies on a passive network that attempts to replace the probe by capacitances to account for the coupling with the sample and impedances to account for stored and dissipated energy. The near-field impedance of the probe due to the presence of a homogeneous sample can then be approximated as:

$$Z_s \approx \frac{1}{j\omega\epsilon_0\epsilon_s v} \quad (10)$$

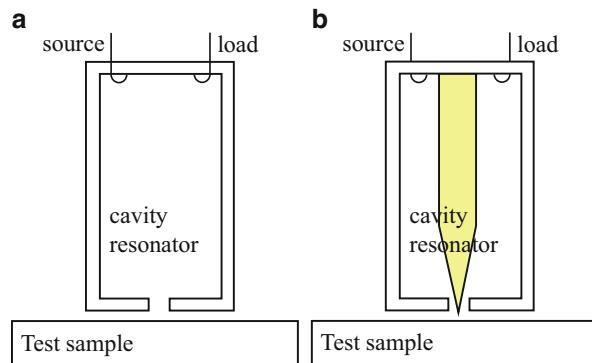
where  $\epsilon_s$  is the complex permittivity of the sample and  $v$  is the volume of the sample in the near field of the probe.

More sensitive instruments use either an aperture in a cavity resonator or a tip in conjunction with a cavity resonator. Figure 11a shows an aperture in a cavity resonator which can be scanned over the sample. Since the cavity is open to the outside, the interaction is through perturbation of the cavity and the resulting shift in the resonant frequency of the cavity. Figure 11b is a modified rectangular resonator (coaxial resonators can also be modified), in which the central conductor extends into a tip outside the cavity to interact with the material. Its operation is similar to that of Fig. 11a, but it is usually more sensitive because of the more localized effect of the tip. In either case, the shift in resonant frequency is due to the perturbation of the cavity. This shift may be written as:

$$\frac{f - f_0}{f} = - \frac{\int_v \Delta\epsilon \mathbf{E} \cdot \mathbf{E}_0^* dv + \int_v \Delta\mu \mathbf{H} \cdot \mathbf{H}_0^* dv}{\int_v \mu \mathbf{H} \cdot \mathbf{H}_0^* dv + \int_v \epsilon \mathbf{E} \cdot \mathbf{E}_0^* dv} \quad (11)$$

**Fig. 11** Resonant microwave microscopy probes.

- (a) Aperture in cavity probe.  
 (b) Tip in aperture probe



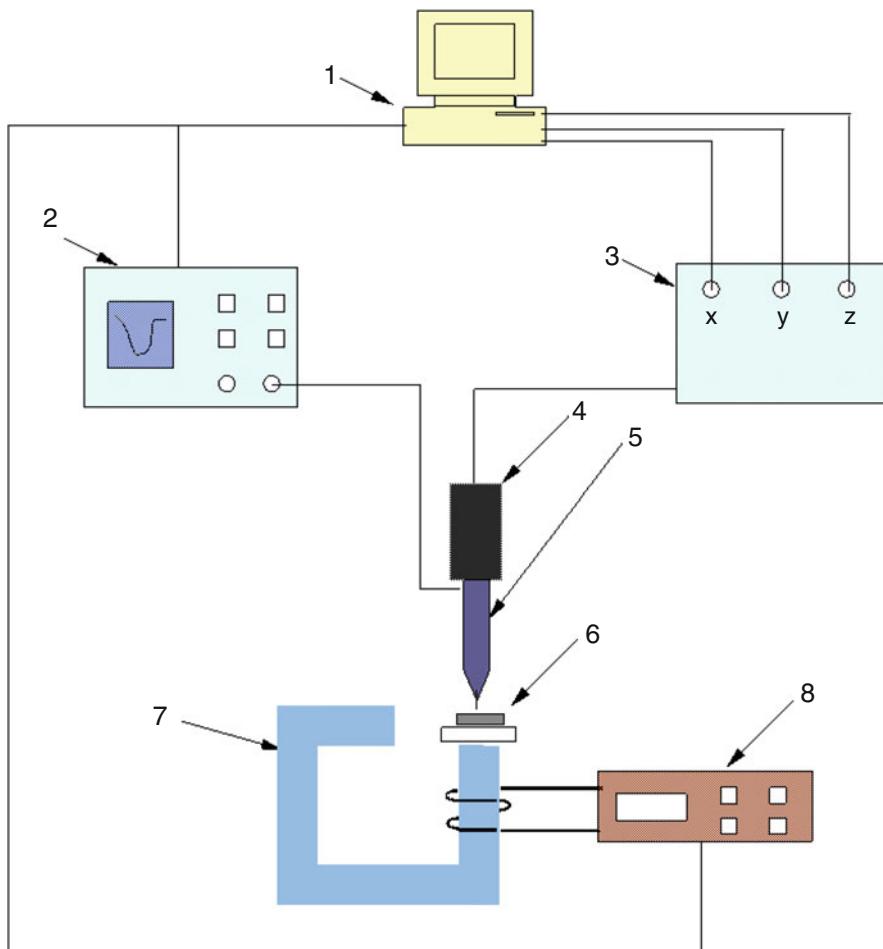
where  $\mathbf{E}$  and  $\mathbf{H}$  are the perturbed fields,  $\mathbf{E}_0$  and  $\mathbf{H}_0$  are the unperturbed fields, and  $\Delta\epsilon$  and  $\Delta\mu$  represent the equivalent change in cavity permittivity and permeability due to the perturbation by the external effects. Although this is a simplistic explanation, it is useful in understanding the shift in resonant frequency of the cavity. In specific cases this can be simplified considerably. For example, in thin films with relative permeability of 1, the shift in resonant frequency can be approximated as (Anlage et al. 2007; Ida 2018):

$$\frac{f - f_0}{f} \approx -(\epsilon_{rf} - 1) \frac{\epsilon_0 \int_{v_f} \Delta\epsilon \mathbf{E} \cdot \mathbf{E}_0^* dv}{\int_v \mu_0 H_0^2 dv + \int_v \epsilon_0 E_0^2 dv} \quad (12)$$

where  $v_f$  is the volume of the film under the influence of the aperture,  $\epsilon_{rf}$  is the relative permittivity of the film, and the denominator is the total energy stored in the cavity. Similar expressions can be obtained for other conditions, but it should be noted that these relations could only be used if the fields and volumes can be identified. In practice they serve to understand the behavior rather than to calculate exact resonant frequencies. If calculations are called for, a full-wave simulation is usually necessary.

The first attempts in near-field microwave microscopy came from the Physics community in attempts to characterize permittivity of thin film materials using ferromagnetic resonance (FMR) (Frait et al. 1960) based on even earlier ideas (Synge 1928). The first microwave probe capable of measuring the spatial variation of magnetic properties was proposed in 1962 (Soohoo 1962). The possibility to perform microwave measurements on thin ferromagnetic layers in a magnetic field was demonstrated more recently (Anlage et al. 2007; Acher et al. 1996; Ustinov et al., 1999). Other applications followed, and recently, the method has found applications in NDE (Ciocan 2000; Zoughi 2000; Ida 1992). A thorough review and analysis of the general method can be found in Anlage et al. (2007).

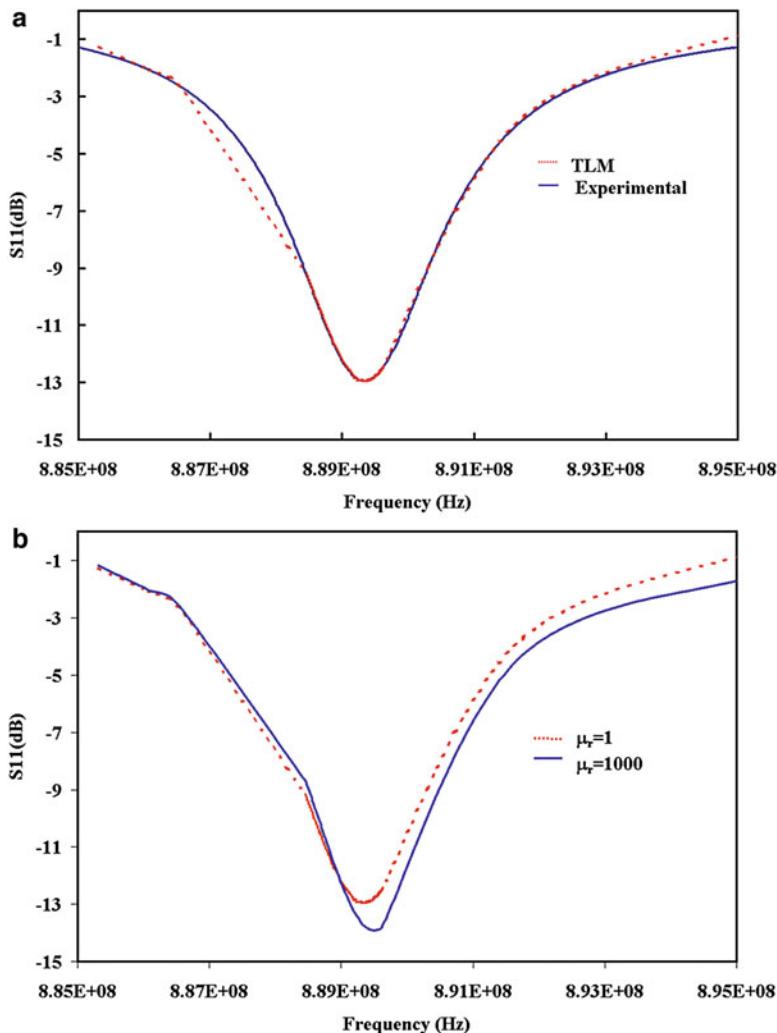
To demonstrate the possibilities in the use of microwave microscopy, consider the simple experimental setup shown in Fig. 12 (Ciocan 2000; Ciocan et al. 2004). In this setup, the magnetic sample (6) was placed over an electromagnet (7) that was



**Fig. 12** Block diagram of the experimental setup: 1, computer; 2, network analyzer; 3, stepper motor controller; 4, port-probe assembly; 5, microwave probe; 6, sample; 7, electromagnet; 8, voltage source. (Reproduced with permission from Ciocan and Ida 2004. © 2004 IEEE)

energized by a dc current from voltage source (8). The sample and the electromagnet assembly were located underneath the microwave probe (5). The personal computer (1) controlled the movement of the microwave probe via a stepper motor controller (3). The network analyzer (2) and voltage source were controlled by the computer. The electromagnet used in this work produced 250 Gauss, and its  $B$  versus  $I$  characteristics were approximately linear in that range as determined with a Gauss meter.

The microwave probe operated in reflection mode (Ida 1992) whereby its  $S_{11}$  parameter was measured by the network analyzer to determine the frequency response of the probe as a function of position over the sample. The probe position over the sample was controlled using micrometers and stepping motors.



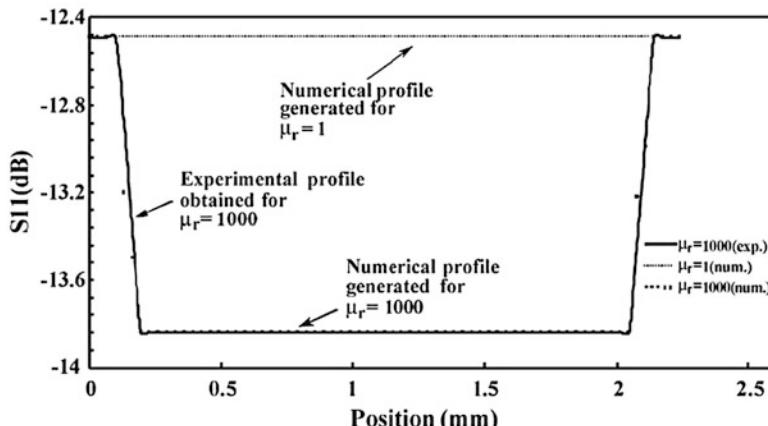
**Fig. 13** (a) Comparison between numerical (dashed line) and experimental frequency responses obtained in the absence of an external magnetic field. (b) Numerical frequency responses obtained in the absence of magnetic field ( $\mu_r = 1$ , dashed line) and in the presence of magnetic field ( $\mu_r = 1000$  solid line). (Reproduced with permission from Ciocan and Ida 2004. © 2004 IEEE)

Figure 13a shows a comparison between numerical and experimental resonance curves obtained for a CO-NETIC alloy sample in the absence of the external magnetic field (Ciocan et al. 2004). The graph shows good agreement between experimental and numerical data, which was obtained with a transmission line method (TLM) (Ciocan and Ida 2004) indicating that simulation of microwave microscopy is a valid endeavor. To further this point, Fig. 3b and Table 1 show simulated responses for the same configuration showing a shift in resonant frequency

**Table 1** Comparison between numerical and experimental data

Type of data	$\mu_r$	Frequency (MHz)	$S_{11}$ (dB)
Numerical	1	889.32	-12.94
Numerical	1000	889.51	-13.91
Experimental	1	889.31	-12.93
Experimental	1000	889.49	-13.90

Table 1 reproduced with permission from Ciocan and Ida 2004. © 2004 IEEE.



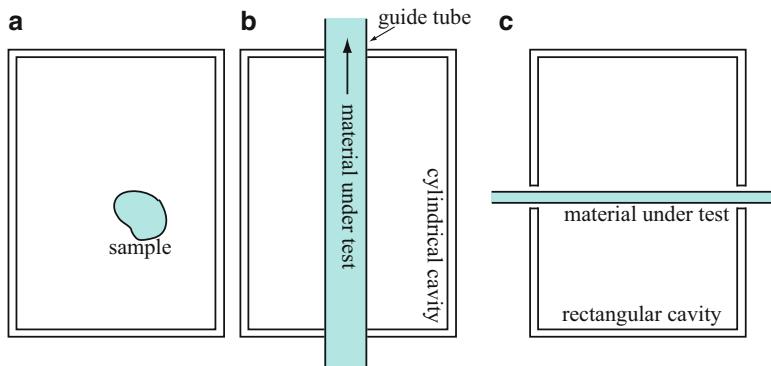
**Fig. 14** Numerical and experimental profiles obtained by scanning over a 2 mm sample. Comparison of simulated and experimental results. (Reproduced with permission from Ciocan and Ida 2004. © 2004 IEEE)

of approximately 190 kHz and a 1 dB difference between the tests with the sample present and absent. The experimental data in Table 1 shows very close values – 180 kHz and 1 dB.

To show that the system can detect nonuniformity in magnetization, the values of the  $S_{11}$  parameter obtained at 889.61 MHz, when the microwave probe was scanned over a 2 mm sample, are plotted in Fig. 14. The comparison with experimental measurements shows indistinguishable values, an indication of the accuracy of the method. It should however be noted that the scan is over a 2 mm sample. Larger samples can be used, but the scan can take considerable time to perform since it is assumed that measurements are taken with the probe stationary or moving very slowly to avoid distortions in the fields.

## Resonant Testing Methods

Resonant methods of testing offer, perhaps, the highest sensitivity of all microwave methods as was also discussed in the previous section. Because what is usually measured is the resonant frequency and the quality factor of the cavity, and the



**Fig. 15** Testing in resonant cavities. (a) A small sample introduced into the cavity. (b) A tubular material flowing through a cylindrical cavity. (c) Testing of flat sample in a partially open rectangular cavity

quality factor itself can be obtained from frequency measurements, resonant methods of testing can also be quantitatively accurate and mostly noise free. With the use of a network analyzer as the basic instrument in the test, high-resolution tests are the norm. The classical approach is to introduce the test sample in the cavity and measure the shift in resonant frequency due to the sample dimensions and its electric properties from which one or more parameters may be monitored. This is shown in Fig. 15a. That of course is also the limitation of resonant methods – what is being measured are the effective properties of the cavity, and it is rather difficult to distinguish between a flaw in the material and a change in dimensions of the sample from measurements of frequency alone. For that reason, resonant methods should be viewed as measuring the real and imaginary part of permittivity (dielectric constant and loss tangent) and/or the real and imaginary part of permeability. Anything that can be correlated to these properties can then be the subject of testing. One can envision testing for diverse properties from dimensional changes to density, but the most common use is in monitoring of moisture content and properties that can be related to moisture content such as curing or drying of substances.

Of course, the configuration in Fig. 15a is limited to small samples that can fit in a cavity. It also suffers from limitations in access to the cavity and positioning of the sample in the cavity reducing its usefulness as a practical test method to a limited number of applications. However, cavities do not have to be closed entirely. Figure 15b and c show possible ways by which test material may be introduced into cavities either for testing of samples or for continuous monitoring on a production line. In Fig. 15b the material is introduced through appropriate openings in the cavity, such as a thin dielectric tube to allow movement of the material in the cavity. The cavity shown is cylindrical, assuming the material under test is itself cylindrical, but other arrangements and cavities are possible. In Fig. 15c, the cavity is split in two forming a small gap between the lower and upper halves to allow insertion and movement of the test material, in this case thin flat samples such as sheet products

(wood, plastics, or fabrics). Resonators may also be made with transmission lines so that they are open (Ida 2015, 2018; Pozar 1998) in many useful configurations. However, the configuration in Fig. 15a is the simplest to analyze using the perturbation method mentioned above and hence affords understanding of what one can expect from resonant methods. Eq. (11) shows the shift in resonant frequency of a cavity if the permeability and permittivity in the cavity change by  $\Delta\mu$  and  $\Delta\varepsilon$  assuming the change occurs throughout the cavity (Fig. 16b). This would be the case, for example, where the cavity is used to sense a gas or monitor humidity. The permittivity in Eq. (10) may be replaced with a complex permittivity  $\varepsilon_c$  to take into account lossy materials:

$$\varepsilon_c = \varepsilon' + j\varepsilon'' = \varepsilon' - j\frac{\sigma_s}{\omega} \quad (13)$$

Equation (10) now becomes (Ida 2018):

$$\frac{f - f_0}{f_0} = -\frac{\int_v \left[ \left( \Delta\varepsilon' - j\frac{\Delta\sigma}{\omega} \right) \mathbf{E} \cdot \mathbf{E}_0 - \Delta\mu \mathbf{H} \cdot \mathbf{H}_0 \right] dv}{\int_v [\varepsilon_0 \mathbf{E} \cdot \mathbf{E}_0 - \mu_0 \mathbf{H} \cdot \mathbf{H}_0] dv} \quad (14)$$

where, again, the change in permeability may be null or it may be complex. In the denominator,  $\varepsilon_0$  and  $\mu_0$  are the properties of the empty cavity.

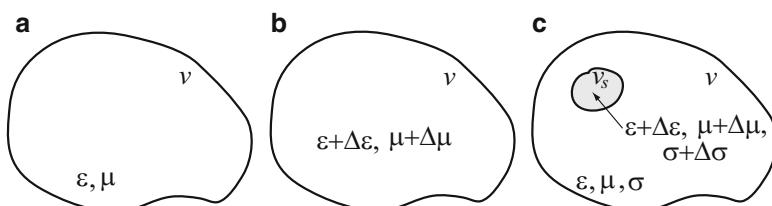
The quality factor may be equally computed from its definition as the peak stored energy divided by the energy dissipated per period:

$$Q = 2\pi \frac{\text{Peak stored energy}}{\text{Energy dissipated}/\text{per period}} \quad (15)$$

Assuming power is lost only in the dielectric, the Q-factor is:

$$Q = \frac{\int_v \varepsilon_0 E_0^2 dv}{\int_v [\varepsilon'' \mathbf{E} \cdot \mathbf{E}_0] dv} \quad (16)$$

The Q-factor may in fact be measured by measuring the bandwidth of the cavity:



**Fig. 16** The concept of cavity perturbation. (a) Unperturbed cavity. (b) Whole cavity perturbation, lossless material. (c) Perturbation by small sample of lossy material

$$Q = \frac{2\pi f_0}{BW} = \frac{f_0}{f_u - f_l} \quad (17)$$

where  $f_u$  is the upper half power frequency and  $f_l$  is the lower half power frequency of the cavity. Clearly then all that is necessary is to measure the frequency response of the loaded cavity w and empty cavity  $f_0$  to obtain the shift in resonant frequency and the  $Q$ -factor. Measurements of the shift in resonant frequency ( $f - f_0$ ) then provide a measure of the change in material properties, whereas measurement of the  $Q$ -factor is a measure of losses in the cavity. From these one may infer the change in properties due to introduction of the material in the cavity.

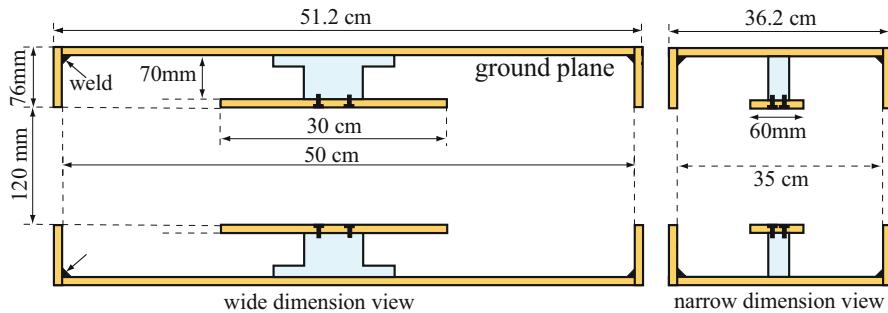
So far it was assumed that the sample fills the cavity. This is not necessarily the case as can be seen in Fig. 15b and c. If the sample only fills a small section of the cavity (see Fig. 16c), the integral in the nominator in Eq. (14) and in the denominator in Eq. (16) only extends over the volume of the small sample. Although calculations using these equations are obviously difficult, from a testing point of view, they show what changes as material properties change.

## Testing in Microwave Cavities

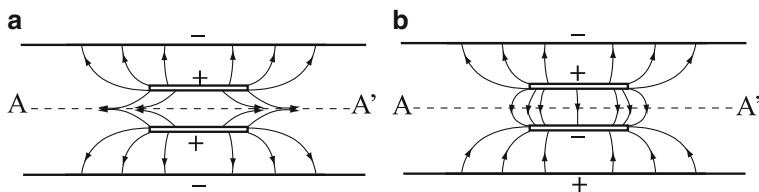
The testing configuration in Fig. 16a is perhaps not as common as other microwave methods, but it can afford as simple method of evaluation of material properties of samples of materials as these depend on frequency, especially when these materials are lossy, mixtures, anisotropic, or nonlinear and a bulk equivalent permittivity is needed, perhaps as a means of verification of other tests. To see the utility of this method, the following shows some simulations of lossy samples in cavities.

## Monitoring of Moisture

Because the resonant frequency of the cavity is due to the whole volume of the cavity, the method is particularly useful in testing and monitoring of material parameters rather than, say, flaws or individual inclusions. The latter will also change the resonant frequency, but localization of defects cannot be done directly, and flaws cannot be distinguished from changes in material properties unless additional testing is done, often using other methods. However, the method is particularly useful in detecting or monitoring changes in material properties over the whole sample or over the whole volume of the cavity. Consider as an example the monitoring of moisture content in a fabric coated with latex on the production line (producing tire belts). The fabric is nylon, polyester, or aramid and is coated with latex in a water solution. The purpose of monitoring is twofold. First it monitors the drying process. More importantly it monitors the amount of solids left as a coating on the fabric after drying to ensure proper coating and hence the performance of the fabric in tires. In the application described here, the resonant sensor was used to control the amount of latex during production. The resonant sensor is shown in Fig. 17. It consists of two



**Fig. 17** Final dimensions of the sensor with internal dimensions of 50 cm by 35 cm. The center plates are 30 cm by 6 cm and are flush with the shield. The separation between the two halves is 120 mm to allow ample space for the fabric. The two probes are used for reflection and transmission measurements. (Reproduced with permission from Open Resonator Microwave Sensor System for Industrial Gauging, N. Ida, The Institution of Engineering and Technology, 2018, ISBN 978-1-78561-140-7)



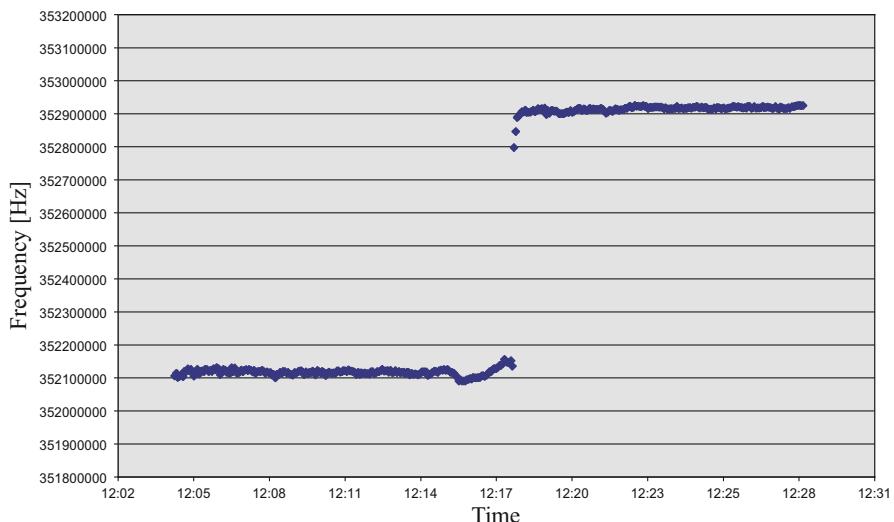
**Fig. 18** Fields in broadside-coupled striplines. (a) Even mode. (b) Odd mode. (Reproduced with permission from Open Resonator Microwave Sensor System for Industrial Gauging, N. Ida, The Institution of Engineering and Technology, 2018, ISBN 978-1-78561-140-7)

open metal shields 50 cm by 35 cm by 7 cm and two center plates 30 cm by 6 cm isolated from the shields using Teflon blocks. Two probes (short antennas) penetrate through the upper shield to couple to the upper central plate. These are connected to a network analyzer to the source and load ports so that both the reflection and transmission properties may be monitored using *S*-parameters. The fabric shown at the center of the sensor is between 0.3 mm and 0.8 mm thick and absorbs up to 35% of a solution of latex in water with the latex being about 20% by volume. The target is 5% solids coating (after drying) on the fabric. The sensor described here is a broadside-coupled stripline resonator (Ida 2018; Garg et al. 2013). Its value in this application is that it is wide open allowing free movement of the fabric and in the fact that it resonates in two modes as shown in Fig. 18. The even mode electric fields are parallel to the fabric and hence sensitive to the properties of the fabric (primarily the solution), whereas the odd mode is more sensitive to the bulk between the center plates rather than the fabric. The even mode is used to monitor the properties of the fabric, whereas the odd mode is used to compensate for bulk effects such as temperature and humidity. Figure 19 shows the upper half of the sensor without

the fabric. The upper center plate and its Teflon block can be seen as well as one of the probes. The sensor is designed so it can move across the fabric to evaluate the whole width of the fabric as it moves. Figure 20 shows the performance of the sensor for two fabrics. The left part is for a fabric with higher permittivity (or more solution



**Fig. 19** The upper half sensor and its attachment to the motion mechanism. Note the construction of the ground plane and partial shield afforded by the bent plates. The second half shell has been moved away for a better view. The Delrin calibration frame can be seen in the lower part of the picture away from the half shell. (Reproduced with permission from Open Resonator Microwave Sensor System for Industrial Gauging, N. Ida, The Institution of Engineering and Technology, 2018, ISBN 978-1-78561-140-7)

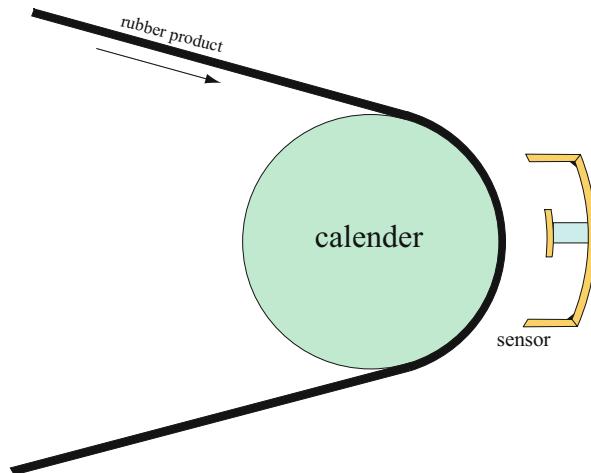


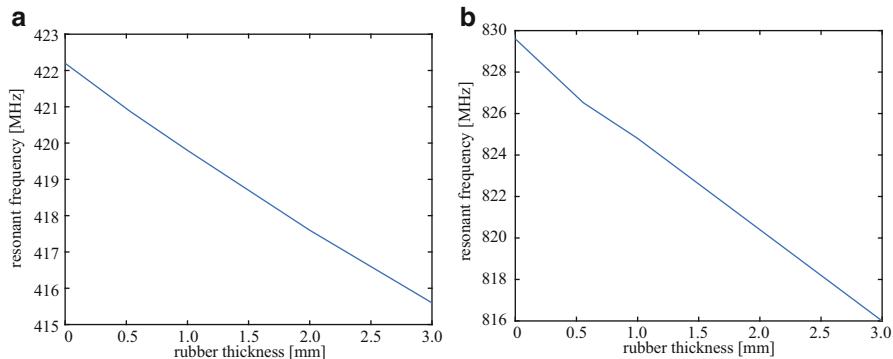
**Fig. 20** Resonant frequency versus time for fabric code (type) and speed change. (Reproduced with permission from Open Resonator Microwave Sensor System for Industrial Gauging, N. Ida, The Institution of Engineering and Technology, 2018, ISBN 978-1-78561-140-7)

on the fabric), whereas the right part is for a lower permittivity fabric (or less solution on the fabric). The sensor is capable of measuring a resolution of 0.014% dry latex on the fabric assuming a 10 kHz resolution at the network analyzer. In principle the resolution can be increased by a factor of 10 (to 0.0014%) for a network analyzer capable of resolving down to 1 kHz, which is not unusual for modern vector network analyzers. Measurements are in real time, that is, the fabric moves continuously, and the sensor moves back and forth across the fabric. Each point in Fig. 20 is the average of five individual measurements of the network analyzer, recorded on a computer (with a time and position stamp). Separately (not shown) the sensor also records temperature and humidity to compensate for errors in the measurements using the odd-mode resonant frequency (Ida 2018; Ida and Bhuyia 2008). The measurements were taken unattended remotely with the operation of the system and transfer of data being done over the Internet.

An example of the use of the same basic sensor for dimensional measurements on rubber is shown in Fig. 21. The rubber is backed by a conductor, and the sensor is made of only one, modified section of the sensor in Fig. 17. The resonator now is between the center plate and the conducting drum, which moves the rubber sheet. The minimum separation between the plate and the drum is 35 mm. The first (lower) and second (higher) resonant frequencies as a function of rubber thickness are shown in Fig. 22, resulting in a frequency shift of 2.4 MHz/mm of rubber at the lower resonant frequency (Fig. 22a). The net effect is a sub-micrometer resolution with a linear response to rubber thickness. At a network analyzer resolution of 1 kHz, the sensor can reliably measure variations in rubber thickness of 0.4 micrometers. Higher resolution can be obtained using the second resonant mode (about twice as high) although in this application the higher resolution was not necessary and work at the lower frequency proved to be easier.

**Fig. 21** Monitoring of rubber thickness on the production line using the open transmission line resonant sensor. (Reproduced with permission from Open Resonator Microwave Sensor System for Industrial Gauging, N. Ida, The Institution of Engineering and Technology, 2018, ISBN 978-1-78561-140-7)





**Fig. 22** Resonant frequency as a function of rubber thickness. (a) For the first resonant frequency. (b) For the second resonant modes

Either implementation of the sensor can be used for other purposes including in situ testing of coatings, concrete, composites, gasses, liquids, and others with similar sensitivities.

---

## Other Methods

### UWB Radar

An ultra-wideband (UWB) radar is characterized by a much wider bandwidth coupled with low power transmission. This is achieved by transmitting a very narrow pulse, typically around 1 ns wide, thus generating a wide-spectrum transmission. Power transmitted is typically under 10 mW. The bandwidth is very wide but typically is taken to be about 25% of the center frequency. As a result of these properties, UWB radars are very compact, intended for relatively short ranges. For NDE purposes, UWB radars offer some exciting possibilities without some of the issues associated with conventional radars. First and foremost, the wideband allows interaction with flaws on the whole spectrum, something that is akin to multi-frequency testing. This means that one can hope to detect and possibly image flaws with much more detail and higher resolution possible with other methods. Second, because of the very low power and its spread over a wide spectrum, there are fewer regulatory constraints and little concern to radiation safety. In addition, there is virtually no concern to interference with or from the system and a typical decrease in the dead zone exhibited by pulsed radars. The main trade-off associated with the wider bandwidth is a reduction in the signal-to-noise ratio. UAB radars can transmit in narrow patterns (high directivity) for increased detectability and resolution through the use of antenna arrays. In spite of what may seem as a complex system, commercial UAB radars are small and highly integrated and often come on a small board with all electronics needed for operation. Single-chip radars are available as

well. The cost is also very low in comparison to most microwave systems. In most NDE applications, the UAB radar is used to scan the object under test and obtain a signal not unlike the A-scans obtained in ultrasound testing. Imaging is much more complex than that and requires both additional equipment and appropriate software. UAB radars come in various center frequencies ranging from a few hundred MHz to the high GHz region. For example, the FCC allows unlicensed use of UWB devices in the frequency range from 3.1 GHz to 10.6 GHz and a maximum equivalent isotropic radiated power (EIRP) spectral density of  $-41.3 \text{ dBm/MHz}$ . Although the band is used for other applications as well, including cell communication and Wi-Fi routing, the fact that UWB applications under these conditions do not interfere with other services allows their use on the same band. Other frequencies and other power levels require licenses.

### Some Applications of UWB Radar

UWB radars are not new – the first were developed in the late 1960s but only recently have they become small enough and inexpensive enough to allow their use in common applications. One of the more common general applications is what is sometimes called “see-through-wall systems” (Yang 2008; Gubinelli et al. 2014; Pochanin et al. 2016). In these, as in GPR one attempts to detect, quantify, and image objects and conditions buried in or located behind lossy dielectrics in a manner similar to GPR, often without making contact with the surface. Examples abound. Gubinelli et al. (2014) used a comparative method to detect surface defects in carbon fiber composites by measuring the mismatch between the reflected waveform from sound sample flawed samples. Cristofani et al. (2017) discuss the use, advantages, and limitations of UWB radar in testing of aeronautic components as well as variations of UWB and the use of synthetic-aperture radar (SAR) algorithms to enhance performance. Imaging and applications to testing of composite materials are also discussed. Xu et al. (2013) show that detection of rebar in concrete using essentially GPR methods in conjunction with UWB and B-scans results in high-resolution images as well as the possibility of estimation of rebar depth in pavements at vehicle road speeds. UWB radar can also be applied on a larger scale in evaluation of conditions of walls and tunnels. Herrmann R (2011) discusses methods and measurements of walls in tunnels in salt mines. 2D and 3D images of conditions as deep as 2 m in the walls show flaws and cracks of various sizes. Other applications can be found in Gubinelli et al. (2014), Zhong et al. (2014), Xie et al. (2006), and Xu et al. (2004). Another promising area of application is in biomaterials and in medicine. The low power and high resolution offered by UWB systems lend themselves naturally to diagnostics in the body and in particular for breast cancer applications (Bidhendi et al. 2014; Fear et al. 2002a, b; Joines et al. 1994). The main impetus for this work is the high contrast between tumors and healthy tissue at microwave frequencies both in terms of permittivity and conductivity. Multiple antennas around the breast allows detection and reconstruction of tumors with minimum discomfort and practically no risk.

UWB radar is rather new in NDE, and many applications are still in the development stage. For that reason, much of the work is done by scanning over the surface

of the test object to produce A-type scans. However, both commercial and experimental devices that incorporate imaging for high-resolution testing have been developed (Shin et al. 2016; Pochanin et al. 2016; Yang et al. 2017; Kidera et al. 2010; Li and Meng 2016).

## Microwave Radiometry

In a radiometric method one measures the electromagnetic radiation and characterizes the distribution of the radiation power by means of a radiometer – a device that measures the radiant flux of the electromagnetic wave. Radiometry is commonly used in optics and in communication (e.g., in the evaluation of antennas and of communication links). In that context, the radiated power is measured using any of a number of sensors such as bolometers, diodes, or even calorimetric methods which respond to the power (or power density) being measured by a rise in their temperature which in turn changes some parameter of the sensor such as resistance or the direct output of thermocouples (Grant 2011; McCluney 2014). These measurements can be correlated with a variety of conditions. It is rather simple, for example, to measure the radiation pattern of an antenna or the power distribution from a distant star. In communication, radiometry is often used to characterize the path of communication and quantify the effects of scatterers such as clouds, dust, or even insect swarms in the path on the communication link. In this type of measurement, one measures the transmitted power through the scatterers (Feliciano 2009; Sweet 2013; Nessel 2015). Radiometry can be active (in which a source of radiation is provided), or it can be passive (in which the natural radiation from objects is measured) (Ulaby et al. 1981; Appleby et al. 2004). In general, passive radiometric methods are used for remote sensing taking advantage of naturally emitted microwave radiation, a method not unlike remote thermal sensing. In NDE however the active radiography is more practical using either a transmission or reflection method. The use of radiometric methods in microwave and mm-wave NDE is but a natural extension of the method to characterize materials and conditions based on the effect they have on the radiated energy. The sample to be tested is irradiated from a source at an appropriate frequency and power density. The wave transmitted through the sample is attenuated by the material, and any discontinuity within the sample scatters the wave. The transmitted wave's power density is measured either by an array of sensors or by a scanning sensor to obtain an image of the transmitted power density (Viegas et al. 2017; Shibuya et al. 2007; Bakhtiari et al. 1997). Alternatively, the wave is reflected off the sample to detect primarily surface and near-surface features (Clancy et al. 2000; Seah et al. 2012; Harmer et al. 2016; Smulders 2012). Typically, the source is broadband (often characterized as “noise”), but there are no specific limitations on the sources other than availability and quality of the waves generated. Nevertheless, higher frequency content is useful since resolution is related to the wavelength, and as in any microwave method, the higher the frequency, the higher the resolution and, often, the signal-to-noise ratio. In the microwave and mm-wave range, this is a particularly difficult issue because sources are not easy to come by

beyond about 100 GHz (free-space wavelength of 3 mm). Nevertheless, applications in the W band (94 GHz) have been reported in NDE of materials and products and in medical applications (Appleby et al. 2004; Clancy et al. 2000; Seah et al. 2012; Harmer et al. 2016; Bardati et al. 1992; Bakhtiari et al. 2012; Shibuya et al. 2007).

## References

- ACGIH Worldwide publication (2017) 2017 threshold limit values (TLV) and biological exposure indices (BEI), published by the American Conference of Governmental Industrial Hygienists (ACGIH)
- Acher O, Baclet P, Perin G (1996) Investigation of the microwave permeability of ferromagnetic film laminations under a magnetic field. *J Magn Magn Mater* 157/158:480–481
- Akuthota B, Hughes D, Zoughi R, Myers J (2004) Near-field microwave detection of disbond in fiber reinforced polymer composites used for strengthening concrete structures and disbond repair verification. *ASCE J Mater Civ Eng* 16:540–546
- Anderson PL (1997) Development of coating thickness instrument for graphite/epoxy composites. In: Review of progress in quantitative nondestructive evaluation, vol 16. Plenum, New York, pp 1601–1607
- Anlage SM, Talanov VV, Schwartz AR (2007) Principles of near-field microwave microscopy. In: Kalinin SV, Gruverman A (eds) Scanning probe microscopy: electrical and electromechanical phenomena at the nanoscale, vol 1. Springer-Verlag, New York, pp 215–253
- Appleby R, Anderton RN, Thomson NH, Jack JW (2004) The design of a real-time 94-GHz passive millimetre-wave imager for helicopter operations. *Passive Millimetre Wave Terahertz Imag Technol* 5619:38–46. <https://doi.org/10.1117/12.581336>
- Arunachalam K, Melapudi VR, Udpa L, Udpa SS (2006) Microwave NDT of cement-based materials using far-field reflection coefficients. *NDT&E Int* 39:585–593
- Ash EA, Nicholls G (1972) Super-resolution aperture scanning microscope. *Nature* 237:510–512
- Auld BA (1978) Theory of ferromagnetic resonance probes for surface cracks in metals. E.L. Ginzton Lab., Stanford University, Stanford. G.L. rep. 2839
- Bahr AJ (1978) Microwave NDE of ceramics. In: Proceedings of the ARPA/AFML review of progress in quantitative NDE, pp 236–241
- Bahr AJ (1981) Microwave eddy-current techniques for quantitative nondestructive evaluation. In: Birnbaum G, Free G (eds) Eddy-current characterization of materials and structures. ASTM STP 722. American Society for Testing and Materials, Philadelphia, pp 311–331
- Bahr AJ (1982) Microwave nondestructive testing methods. Gordon and Breach, Newark
- Bachthari S, Zoughi R (1990) Microwave thickness measurement of lossy layered dielectric slabs using incoherent reflectivity. *Res Nondestruct Eval* 2:157–168
- Bachthari S, Qaddoumi N, Ganchev SI, Zoughi R (1994) Microwave noncontact examination of disbond and thickness variation in stratified composite media. *IEEE Trans Microwave Theory Tech* 42:389–395
- Bakhtiari S, Ganchev S, Zoughi R (1993) Open-ended rectangular waveguide for nondestructive thickness measurement and variation detection of lossy dielectric slabs backed by a conducting plate. *IEEE Trans Instrum Meas* 42:19–24
- Bakhtiari K, Gopalsami N, Raptis C (1997) Determining bonding quality in polymer composites with a millimeter wave sensor. In: Review of progress in quantitative nondestructive evaluation. Springer, New York, pp 665–671
- Bakhtiari S, Elmer T, Cox NM, Gopalsami N, Raptis A, Liao S, Mikhelson I, Sahakian A (2012) Compact millimeter-wave sensor for remote monitoring of vital signs. *IEEE Trans Instrum Meas* 61:830–841
- Balanis CA (2005) Antenna theory, 3rd edn. Wiley Interscience, Hoboken

- Bardati F, Brown VJ, Ross MP, Tognolatti P (1992) Microwave radiometry for medical thermal imaging: theory and experiment. *IEEE MTT-S Int Microwave Symp Dig* 3:1287–1290. <https://doi.org/10.1109/MWSYM>
- Bidhendi HK, Jafari MH, Genov R (2014) Ultra-wideband imaging systems for breast cancer detection. In: Yuce MR (ed) *Ultra-wideband and 60 GHz communications for biomedical applications*. Springer Science+Business Media, New York. [https://doi.org/10.1007/978-1-4614-8896-5\\_5](https://doi.org/10.1007/978-1-4614-8896-5_5)
- Bogosanovic M, Al Anbuky A, Emms GW (2012) Microwave nondestructive testing of wood anisotropy and scatter. *IEEE Sensors J* 13:306–313
- Bois KJ, Benally A, Zoughi R (2000) Microwave near-field reflection property analysis of concrete for material content determination. *IEEE Trans Instrum Meas* 49:49–55
- Caorsi S, Massa A, Pastorino M, Randazzo A (2003) Electromagnetic detection of dielectric scatterers using phaseless synthetic and real data and the memetic algorithm. *IEEE Trans Geosci Remote Sens* 41:2745–2753
- Cassidy D, Holton G, Rutherford J (2002) *Understanding physics*. Springer, New York, p 563
- Chen G, Hu B, Takeuchi I, Chang KS, Xiang XD, Wang G (2005) Quantitative scanning evanescent microwave microscopy and its applications in characterization of functional materials libraries. *Meas Sci Technol* 16:248–260. <https://doi.org/10.1088/0957-0233/16/1/033>
- Ciocan R (2000) Evanescent microwave microscope: a novel nondestructive, high resolution method for quantitative evaluation of materials. MS thesis, Case Western Reserve University, pp 39–46
- Ciocan R, Ida N (2004) Transmission line matrix model for detection of local changes in permeability using a microwave technique. *IEEE Trans Magn* 40:651–654. <https://doi.org/10.1109/TMAG.2004.824883>
- Ciocan R, Ida N, Ciocan E, Jiang H (2004) Applications of the transmission line matrix method to microwave scanning microscopy. *ACES J* 19:94–100
- Clancy T, Ulaby F, Dobson C (2000) Detection of Space Shuttle tile defects using millimeter-wave radiometry. In: *IEEE antennas and propagation society international symposium*, vol 3. Salt Lake City, pp 1787–1790
- Cristofani E, Friederich F, Vandewal M, Jonuscheit J (2017) Nondestructive testing of aeronautics composite structures using ultrawideband radars. In: Taylor J (ed) *Signals, targets, and applications*. CRC Press, Boca Raton, pp 237–256
- Das A, Das SK (2000) *Microwave engineering*. McGraw-Hill, Boston
- De Rosny J, Lerosey G, Fink M (2010) Theory of electromagnetic time-reversal mirrors. *IEEE Trans Antennas Propag* 58:3139–3149
- Decretor M, Gardiol FE (1974) Simple nondestructive method for measurement of complex permittivity. *IEEE Trans Instrum Meas* IM-23:434–438
- Dunsmore JP (2012) *Handbook of microwave component measurements: with advanced VNA techniques*. John Wiley & Sons, New York
- Fear EC, Hagness SC, Meaney PM, Okoniewski M, Stuchly MA (2002a) Enhancing breast tumor detection with near-field imaging. *IEEE Microw Mag* 3:48–56
- Fear EC, Li X, Hagness SC, Stuchly MA (2002b) Confocal microwave imaging for breast cancer detection: localization of tumors in three dimensions. *IEEE Trans Biomed Eng* 49:48–56
- Feliciano W (2009) Ka-band LEO Ground Station Development for radiometry and rain data collection, and rain attenuation analysis. Masters' thesis, The University of Akron, Akron
- Frait Z, Kambersky V, Malek Z, Ondris M (1960) Local variations of uniaxial anisotropy in thin films. *Czech J Phys* 10:616–617
- Ganchev S, Qaddoumi N, Brandenburg D, Bakhtiari S, Zoughi R, Bhattacharyya J (1994) Microwave diagnosis of rubber compounds. *IEEE Trans Microwave Theory Tech* 42:18–24
- Ganchev S, Qaddoumi N, Ranu E, Zoughi R (1995) Microwave detection optimization of disbonds in layered dielectrics with varying thicknesses. *IEEE Trans Instrum Meas* IM-44:326–328
- Garg R, Bahl I, Bozzi M (2013) *Microstrip lines and slotlines*. Artech House, Boston
- Ghasr MT (2004) Detection and size evaluation of corrosion precursor pitting using near-field microwave and millimeter wave nondestructive testing methods. Master's thesis, Electrical and Computer Engineering Department, University of Missouri- Rolla, Rolla

- Ghasr M, Carroll B, Kharkovsky S, Zoughi R, Austin R (2005a) Size evaluation of corrosion precursor pitting using near-field millimeter wave nondestructive testing methods. In: Proceedings of 31st annual review progress in quantitative nondestructive evaluation, vol 760, pp 547–553
- Ghasr M, Kharkovsky S, Zoughi R, Austin R (2005b) Comparison of near-field millimeter wave probes for detecting corrosion precursor pitting under paint. *IEEE Trans Instrum Meas* 54:1497–1504
- Ghasr MT, Horst MJ, Lechuga M, Rapoza R, Renoud CJ, Zoughi R (2015) Accurate one-sided microwave thickness evaluation of lined-fiberglass composites. *IEEE Trans Instrum Meas* 64:2802–2812
- Ghodgaonkar DK, Ali NA (2000) Microwave nondestructive testing of composite materials using free-space microwave measurement techniques. In: Proceedings of 15th world conference on non-destructive testing, pp 15–21
- Ginzton EL (1975) The \$100 idea. *IEEE Spectr* 12:30–39
- Godgaonkar DK, Varadan VV, Varadan VK (1990) Free space measurement of complex permittivity and complex permeability of magnetic materials at microwave frequencies. *IEEE Trans Instrum Meas* 39:387–394
- Grant BG (2011) Field guide to radiometry. SPIE Digital Library. <https://doi.org/10.1117/3.903926>
- Gray S, Zoughi R (1997) Dielectric sheet thickness variation and disbond detection in multi-layered composites using an extremely sensitive microwave approach. *Mater Eval* 55:42–48
- Gubinelli S, Paolini M, Andrea Giorgetti A, Troiani E, Chiani M (2014) An ultra-wideband radar approach to nondestructive testing. In: IEEE International Conference on Ultra-WideBand (ICUWB). <https://doi.org/10.1109/ICUWB.2014.6958997>
- Han J, Nguyen C (2007) Development of a tunable multiband UWB radar sensor and its applications to subsurface sensing. *IEEE Sensors J* 7:51–58
- Harmer SW, Shylo S, Shah M, Bowring NJ, Owda AY (2016) On the feasibility of assessing burn wound healing without removal of dressings using radiometric millimetre-wave sensing. *Prog Electromagn Res M* 45:173–183
- Herrmann R (2011) M-sequence based ultra-wideband radar and its application to crack detection in salt mines. Ph. D dissertation, Ilmenau University of Technology, Department of Electrical Engineering and Information Technology. URN: urn:nbn:de:gbv:ilm1-2011000344
- Hochschild R (1963) Application of microwaves in nondestructive testing. *Nondestr Test* 21:115–120
- Huber C, Abiri H, Ganchev SI, Zoughi R (1997) Modeling of surface hairline-crack detection in metals under coatings using an open-ended rectangular waveguide. *IEEE Trans Microwave Theory Tech* 45:2049–2057
- Hughes D, Zoughi R (2005a) A novel method for determination of dielectric properties of materials using a combined embedded modulated scattering and near-field microwave techniques. Part I—forward model. *IEEE Trans Instrum Meas* 54:2389–2397
- Hughes D, Zoughi R (2005b) A novel method for determination of dielectric properties of materials using a combined embedded modulated scattering and near-field microwave techniques. Part II—dielectric property recalculations. *IEEE Trans Instrum Meas* 54:2398–2401
- Hughes D, Wang N, Case T, Donnell K, Zoughi R, Austin R, Novack M (2001) Microwave nondestructive detection of corrosion under thin paint and primer in aluminum panels. *Subsurf Sens Technol Appl: An Int J* (Special issue on advances applications in microwave and millimeter wave nondestructive evaluation) 2:435–451
- ICNIRP (1998) Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz). *Health Phys* 74:494–522
- Ida N (1992) Microwave NDT. Kluwer Press, Dordrecht/Boston, pp 201–273
- Ida N (2008) Open stripline resonator sensors for rubber properties gauging. In: Proceedings of MOMAG-08, Florianopolis, Brazil, 7–10 September 2008, pp 1077–1083
- Ida N (2015) Engineering electromagnetics. Springer, Cham

- Ida N (2018) Open resonator microwave sensor systems for industrial gauging: a practical design approach. IET, London
- Ida N, Bhuyia O (2008) Design and optimization of an open stripline resonator for rubber thickness gauging. In: Proceedings of the 11th international conference on optimization of electrical and electronic equipment, vol 4. Brasov, pp 97–100
- IEEE (2002) (R2008) Recommended practice for measurements and computations of radio frequency electromagnetic fields with respect to human exposure to such fields (2002),100 kHz–300 GHz, Std.
- IEEE (2006) Standard for safety levels with respect to human exposure to radio frequency electromagnetic fields (2006), 3 kHz to 300 GHz, Std.
- International Commission on Non-Ionizing Radiation Protection (ICNIRP) (2016) Inaccurate official assessment of radiofrequency safety by the advisory group on non-ionising radiation. *Rev Environ Health* 31:493–503
- ITU. K.52 (2016) Guidance on complying with limits for human exposure to electromagnetic fields, Std.
- Joffe R, Shavit R, Kamenetskii E (2017) Multiresonance measurement method for microwave microscopy. *IEEE Trans Instrum Meas* 66:2174–2180
- Joines WT, Zhang Y, Li C, Jirtle RL (1994) The measured electrical properties of normal and malignant human tissues from 50 to 900 MHz. *Med Phys* 21:547–550
- Joongsuk, Nguyen C (2005) An ultrawide-band microwave radar sensor for nondestructive evaluation of pavement subsurface. *IEEE Sensors J* 5:942–949
- Jundi AJ, Qaddoumi NN (2009) Open-ended rectangular waveguide near-field frequency response of multilayered structures. In: Proceedings of the 5th EUCAP conference (European conference on antennas and propagation), pp 1304–1308
- Kadaba PK (1984) Simultaneous measurement of complex permittivity and permeability in the millimeter region by a frequency domain technique. *IEEE Trans Instrum Meas* 33:336–340
- Kharkovsky S, Akay MF, Hasar UC, Atis CD (2002) Measurement and monitoring of microwave reflection and transmission properties of cement-based specimens. *IEEE Trans Instrum Meas* 51:1210–1218
- Kharkovsky S, Ryley AC, Stephen V, Zoughi R (2006) Dualpolarized microwave near-field reflectometer for noninvasive inspection of carbon fiber reinforced polymer (CFRP) strengthened structures. In: Proceedings of the IEEE instrumentation and measurement technology conference, Sorrento, pp 2108–2111
- Kharkovsky S, McClanahan A, Zoughi R, Palmer DD (2011) Microwave dielectric-loaded rectangular waveguide resonator for depth evaluation of shallow flaws in metals. *IEEE Trans Instrum Meas* 60:3923–3930
- Kidera S, Sakamoto T, Sato T (2010) Accurate UWB radar three-dimensional imaging algorithm for a complex boundary without range point connections. *IEEE Trans Geosci Remote Sens* 48:1993–2004
- Leonard JD, Stropki CT (1961) Utilization of microwave frequencies for quality control and nondestructive testing of dielectric components. In: Proceedings of 2nd symposium on NDT of aircraft and missile components, San Antonio
- Lerosey G, De Rosny J, Tourin A, Derode A, Montaldo G, Fink M (2004) Time reversal of electromagnetic waves. *Phys Rev Lett* 92:193904
- Li Z, Meng Z (2016) A review of the radio frequency non-destructive testing for carbon-fibre composites. *Meas Sci Rev* 16:68–76. <https://doi.org/10.1515/msr-2016-0010>
- Li Y, Bowler N, Johnson DB (2011) A resonant microwave patch sensor for detection of layer thickness or permittivity variations in multilayered dielectric structures. *IEEE Sensors J* 11:5–15
- Liskow CL (1948) Ultra-high frequency apparatus for inspection of sheet and other materials. US patent 2,457–895
- Liu D, Kang G, Li L, Chen Y, Vasudevan S, Joines W (2005) Electromagnetic time-reversal imaging of a target in a cluttered environment. *IEEE Trans Antennas Propag* 53:3058–3066

- Mast DB (2001) Evanescent microwave sensor scanning for detection of subcoating corrosion. *J Corros Sci Eng* 3, Paper 15, 2001 [Online]. Available: <http://www.jcse.org/Volume2/Paper15/v3p15.html>
- Mazlumi F, Sadeghi SHH, Moini R (2006) Interaction of an open-ended rectangular waveguide probe with an arbitraryshape surface crack in a lossy conductor. *IEEE Trans Microwave Theory Tech* 54:3706–3711
- McCluney WR (2014) Introduction to radiometry and photometry, 2nd edn. Artech House, Boston
- Mehta P, Chand K, Narayanswamy N, Beutner DJ, Zoughi R, Stoeker WV (2006) Microwave reflectometry as a novel diagnostic tool for detection of skin cancers. *IEEE Trans Instrum Meas* 55:1309–1316
- Mubarak K, Bois KJ, Zoughi R (2001) A simple, robust and on-site microwave technique for determining water-to-cement (w/c) ratio of fresh Portland cement-based materials. *IEEE Trans Instrum Meas* 50:1255–1263
- Mukherjee S, Udupa L, Udupa S, Rothwell E (2016) Target localization using microwave time reversal mirror in reflection mode. *IEEE Trans Antennas Propag* 65:820–828
- Mukherjee S, Tamburrino A, Haq M, Udupa S, Udupa L (2018) Far field microwave NDE of composite structures using time reversal mirror. *NDT&E Int* 93:7–17
- Nessel J (2015) Atmospheric phase compensation via multi-spectral decorrelation of water vapor radiometer data. PhD Dissertation, The University of Akron, Akron
- Nicholson AM, Ross GF (1970) Measurement of the intrinsic properties of materials by time-domain technique. *IEEE Trans Instrum Meas* IM-19:377–382
- OET Bulletin 65 (1997) Evaluating compliance with FCC guidelines for human exposure to radiofrequency electromagnetic fields. <https://www.fcc.gov/general/oet-bulletins-line>. Accessed 20 Dec, 2018
- Peer S, Case JT, Gallaher E, Kurtis KE, Zoughi R (2003) Microwave reflection and dielectric properties of mortar subjected to compression force and cyclically exposed to water and sodium chloride solution. *IEEE Trans Instrum Meas* 52:111–118
- Pochanin G, Masalov S, Pochanina I, Capineri L, Falorni P, Bechtel T (2016) Modern trends in development and application of the UWB radar systems. In: 8th international conference on ultrawideband and ultrashort impulse signals (UWBUSIS). <https://doi.org/10.1109/UWBUSIS.2016.7724141>
- Pozar DM (1998) Microwave engineering, 2nd edn. John Wiley & Sons, Inc., New York
- Qaddoumi N, Zoughi R, Carriveau GW (1996) Microwave detection and depth determination of disbands in lowpermittivity and low-loss thick sandwich composites. *Res Nondestruct Eval* 8:51–63
- Qaddoumi N, Shroyer A, Zoughi R (1997) Microwave detection of rust under paint and composite laminates. *Res Nondestruct Eval* 9:201–212
- Qaddoumi NN, Handjojo L, Bigelow T, Easter J, Bray A, Zoughi R (2000) Microwave corrosion detection using open-ended rectangular waveguide sensors. *Mater Eval* 58:178–184
- Qaddoumi N, Bigelow T, Zoughi R, Brown L, Novack M (2002) Reduction of sensitivity to surface roughness and slight standoff distance variations in microwave inspection of thick composite structures. *Mater Eval* 60:165–170
- Qaddoumi NN, Saleh WM, Abou-Khousa M (2007) Innovative near-field microwave nondestructive testing of corroded metallic structures utilizing open-ended rectangular waveguide probes. *IEEE Trans Instrum Meas* 56:1961–1966
- Rayleigh L (1897) On the passage of electric waves through tubes. *Philos Mag* 43:125–132
- Reyes-Rodríguez S, Lei N, Crowgey B, Udupa L, Udupa S (2014) Time reversal and microwave techniques for solving inverse problem in non-destructive evaluation. *NDT&E Int* 62:106–114
- Rohde & Schwarz USA, Inc (2007) Fundamentals of vector network analysis, version 1.1, published by Rohde & Schwarz USA, Inc
- Rosner BT, Van der Weide DW (2002) High-frequency near field microscopy. *Rev Sci Instrum* 73:2505–2525

- Sayar M, Ogawa DSK (2009) Non destructive microwave detection of layer thickness in degraded thermal barrier coatings using K-and W-band frequency range. *NDT&E Int* 42:398–403
- SCENIHR (2015) Potential health effects of exposure to electromagnetic fields (emf). European Commission: DG Health and Food Safety, Luxembourg
- Seah KH, Karim MF, Ong LC, Chiam TM (2012) Rapid corrosion detection using 94 GHz millimeter wave technology. In: IEEE international instrumentation and measurement technology conference (I2MTC), Graz, pp 473–476
- Shibuya T, Goto Y, Dobroiu A, Otani C, Kawase K (2007) Nondestructive detection of carbon in soot collection filters by using a 94 GHz source. In: Joint 32nd international conference on infrared and millimeter waves and the 15th international conference on terahertz electronics, Cardiff, pp 273–274
- Shin HJ, Narayanan RM, Asmuth MA, Rangaswamy M (2016) Ultrawideband noise radar tomography: principles, simulation, and experimental validation. *Int J Microwave Sci Technol* 2016. <https://doi.org/10.1155/2016/5787895>
- Smulders PFM (2012) Analysis of human skin tissue by millimeter-wave reflectometry. *Skin Res Technol* 19:209–216. <https://doi.org/10.1111/j.1600-0846.2012.00629.x>
- Soohoo RF (1962) A microwave magnetic microscope. *J Appl Phys* 33:1276
- Stuchly M, Stuchly S (1980) Coaxial line reflection methods for measuring dielectric properties of biological substances at radio and microwave frequencies—a review. *IEEE Trans Instrum Meas IM-29:176–183*
- Stutzman WL, Thiele GA (2013) Antenna theory and design, 3rd edn. Wiley, Hoboken/New York
- Sweet A (2013) Microwave transmission measurements using a radiometer system. Masters' thesis, The University of Akron, Akron
- Synge EA (1928) A suggested method for extending microscopic resolution into the ultra-microscopic region. *Phil Mag* 6:356–362
- Tabib-Azar M, Garcia-Valenzuela A, Ponchak G (2002) Evanescent microwave microscopy for high resolution characterization of materials. Kluwer, Norwell
- Trabelsi S, Kraszewski A, Nelson S (1997) Simultaneous determination of density and water content of particulate materials by microwave sensors. *Electron Lett* 33:874–876
- Travassos L (2007) Modélisation numérique pour l'évaluation non destructive électromagnétique: application au contrôle non destructif des structures en béton. PhD dissertation, École Centrale de Lyon, France
- Ulaby FT, Moore RK, Fung AF (1981) Microwave remote sensing – active and passive – microwave remote sensing fundamentals and radiometry, vol 1. Addison-Wesley, Reading, pp 344–418
- Ustinov VV, Rinkevich AB, Romashev LN (1999) Microwave magnetoresistance of Fe/Cr multilayers in current-perpendicular to plane geometry. *J Magn Magn Mater* 198/199:82–84
- Viegas C, Alderman B, Huggard PG, Powell J, Parow-Souchon K, Firdaus M, Liu H, Duff CI, Sloan R (2017) Active millimeter-wave radiometry for nondestructive testing/evaluation of composites—glass fiber reinforced polymer. *IEEE Trans Microwave Theory Tech* 65:641–650
- Wang Y, Zoughi R (2000) Interaction of surface cracks in metals with open-ended coaxial probes at microwave frequencies. *Mater Eval* 58:1228–1234
- Weir W (1974) Automatic measurement of complex dielectric constant and permeability at microwave frequencies. *Proc IEEE* 62(1):33–36
- WHO (2016) Who research agenda for radiofrequency fields. WHO, Geneva
- Winters DW, Bond EJ, Van Veen BD, Hagness SC (2006) Estimation of the frequency-dependent average dielectric properties of breast tissue using a time-domain inverse scattering technique. *IEEE Trans Antennas Propag* 54:3517–3528
- Wu Z (2007) Software VNA and microwave network design and characterization. Wiley, New York
- Xie Y, Guo B, Xu L, Li J, Stoica P (2006) Multistatic adaptive microwave imaging for early breast cancer detection. *IEEE Trans Biomed Eng* 53:1647–1657
- Xu L, Davis S, Hagness SC, Weide D, Veen B (2004) Microwave imaging via space–time beamforming: experimental investigation of tumor detection in multilayer breast phantoms. *IEEE Trans Microwave Theory Tech* 52:1856–1865

- Xu X, Xia T, Venkatachalam A, Huston D, ASCE M (2013) Development of high-speed Ultra-wideband ground-penetrating radar for rebar detection. *J Eng Mech* 139:272–285
- Yang Y (2008) Development of a real-time ultra-wideband see through wall imaging radar system. PhD dissertation, University of Tennessee, Knoxville
- Yang X, Zheng YR, Ghasr MT, Donnell KM (2017) Microwave imaging from sparse measurements for near-field synthetic aperture radar. *IEEE Trans Instrum Meas* 66:2680–2692
- Zhong CH, Croxford AJ, Wilcox PD (2014) Remote inspection system for impact damage in large composite structure. *Proc R Soc A Math Phy Eng Sci* 471:20140631
- Zoughi R (2000) Microwave non-destructive testing and evaluation principles. Springer Science and Business Media, New York
- Zoughi R, Gallion JR, Ghasr MT (2016) Accurate microwave measurement of coating thickness on carbon composite substrates. *IEEE Trans Instrum Meas* 65:951–953



# Terahertz Techniques in NDE

27

Joachim Jonuscheit

## Contents

Introduction .....	968
Terahertz Time-Domain Spectrometer .....	969
Terahertz FMCW System .....	971
Imaging Concepts .....	973
Current Applications .....	977
Outlook .....	982
References .....	983

### Abstract

So far, terahertz measurement technology has rarely been used in industry. This book chapter presents therefore the underlying technologies, the resulting possibilities and exemplary applications.

The introduction describes the basic properties of terahertz systems and the advantages of terahertz technology over established techniques such as infrared, X-rays and ultrasound.

The two industrially relevant terahertz techniques, time domain spectrometer and FMCW system, are described in detail. In addition, the metrological basics for their main applications are introduced.

Imaging is required in many applications. Different approaches are presented.

Before the chapter concludes with an outlook on future applications, current application examples are discussed. This concerns applications in the fields of coating thickness measurement, safety, materials research, testing of lightweight materials and inline testing.

An extensive reference list provides a deeper insight into the topics addressed.

---

J. Jonuscheit (✉)

Fraunhofer Institute for Industrial Mathematics ITWM, Kaiserslautern, Germany  
e-mail: [joachim.jonuscheit@itwm.fraunhofer.de](mailto:joachim.jonuscheit@itwm.fraunhofer.de)

## Introduction

Terahertz radiation consists of electromagnetic waves at frequencies from 0.1 to 10 terahertz (THz) or the range of the spectrum between microwaves and infrared light. This frequency band has been hardly used outside of the scientific community. The reason? The methods for generating and detecting terahertz radiation were simply too complex and expensive. It was only at the beginning of the last decade that terahertz systems began to make rapid advances in development: the leap from the lab to the first application was not far off. Terahertz systems are fascinating because the radiation is able to penetrate materials that absorb visible and infrared light. Because many molecules have a spectral fingerprint in this frequency band, the identification of various materials is possible. In addition, large bandwidths can be realized at such high frequencies, which make wireless communications at previously unattainable data rates appear feasible. After many years of research and development, this so-called terahertz gap is closed, and a huge variety of terahertz products are commercially available. The first industrial applications have now been implemented, and they demonstrate that terahertz technology has successfully made the leap from the lab to practical application.

In essence, two different technologies are available for use in today's industrial environments.

One of these includes the terahertz pulsed imaging systems based on time-domain spectroscopy (TDS). The use of ultrashort-pulse lasers generates very short terahertz pulses that are measured sensitively (Bründermann et al. 2012; Mittelman 2003; Naftaly 2015; Theuer et al. 2011). In this manner, very accurate time-of-flight analyses can be performed, for example, to measure thickness or for material identification via spectroscopic studies.

The other technology focuses on frequency-modulated continuous wave (FMCW) electronic terahertz systems that rely either on waveguides (Friederich et al. 2011; Skolnik 2008) or on integrated circuits (ICs) (Jaeschke et al. 2013; Lisauskas et al. 2009; Öjefors et al. 2009). Generally, these electronic systems are based on microwave oscillators, where the emission frequency is increased into the terahertz range by external frequency multipliers. The waveguide-based systems follow a discrete and modular construction, whereas the IC-based systems are highly integrated and combine as many components as possible on a single or, in some cases, a few chips. By integrating the greatest number of high-frequency circuits on one chip, small sizes and cost-effective mass production become possible. The waveguide systems rely on high-frequency circuits, often based on semiconductors in InP or GaAs technology (Tessmann et al. 2014), able to achieve up to approximately 2 THz with current components. Most of the highly integrated IC circuits are based on SiGe technology, which currently enables up to approximately 300 GHz to be realized. ICs up to approximately 100 GHz are already commercially available. However, compared to the waveguide-based systems, the bandwidths achieved up to now are obviously lower and, therefore, less capable for depth resolution.

These variants are at different stages of development and each has its predestined fields of application. Costs are also very different in each technology: the pure

electronic systems are usually the more reasonable of the two, although the costs for the optical systems are sharply dropping. Depending on the system design, terahertz radiation can be used for measurements in transmission, reflection, or scatter patterns. In addition to pure material characterization, these systems are also suitable for imaging and tomographic methods. Terahertz technology has significant advantages over competing methods that rely on microwave, infrared (IR), X-ray, or ultrasonic systems:

- **Versus Microwaves:** Terahertz technology can achieve significantly better spatial resolutions because of the shorter wavelengths. Also, in the terahertz range, spectroscopic fingerprints are available for substance identification.
- **Versus IR:** Terahertz radiation penetrates dielectric materials and has a relatively low Mie and Rayleigh scattering.
- **Versus X-ray:** Terahertz radiation poses no health hazards because of the low photon energy.
- **Versus Ultrasonic:** Terahertz radiation enables the noncontact inspection of materials.

Of course, disadvantages also exist, some of which have been underestimated. For example, the absorption of terahertz waves by water molecules means that results can be affected by moisture in the air. Metal and other electrical conductors reflect the complete terahertz electromagnetic wave and so are not transparent to millimeter and terahertz waves.

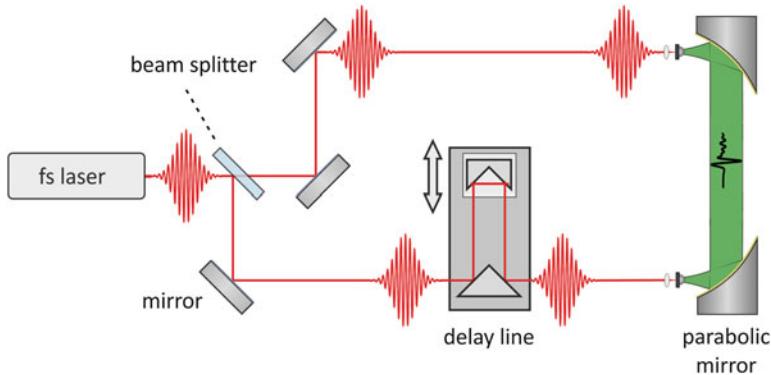
---

## Terahertz Time-Domain Spectrometer

The principle of terahertz time-domain spectroscopy (TDS) is based on the generation and coherent detection of pulsed terahertz radiation using components gated by femtosecond laser pulses. A frequency of 1 THz corresponds to a cycle duration of 1 ps, which means that terahertz frequencies are located at the very limits of the frequency response of electronic devices. With typical commercial fiber lasers (100 fs pulse duration and less), frequencies between 100 GHz and 5 THz can be detected.

Since no electronic device or amplifier can so far directly respond to the high-frequency oscillations of the terahertz electric field, there is a need to reduce the required high-frequency response of the detection scheme (Fig. 1). Different techniques have been realized to detect the fast oscillation. All techniques have in common that they split the laser into two fractions, one for the transmitter and one for the receiver arm. Comparable to a pump-probe experiment, only the relative delay between generation and detection laser pulses has to be taken into account. This relative delay can be realized either mechanically or electronically or optically (Dietz et al. 2014; Kolano et al. 2018; Klatt et al. 2009; Wilk et al. 2011).

In contrast to intensity measurements (e.g., power meters or standard FTIR spectroscopy), terahertz TDS allows recording of the electric field  $E$  directly and not just the intensity. An intensity measurement  $I \propto |E|^2$  would lose the entire phase information. Applying terahertz TDS, the signed digitized electric field  $E(t)$  is



**Fig. 1** Sketch of a terahertz time-domain spectroscopy system

sampled in the time domain as a function of the delay  $t$ . A Fourier transformation connects the quantity in the time domain to an equivalent complex quantity in the frequency domain  $\tilde{E}(\omega)$ .

Consider a broadband THz pulse  $\tilde{E}(\omega)_{\text{in}}$  in propagating through a medium of length  $l$ , and then the amplitude and the phase of the pulse will be changed

$$\tilde{E}(\omega)_{\text{out}} = \tilde{E}(\omega)_{\text{in}} e^{-i\tilde{k}}(\omega)l$$

with  $\tilde{k}$  the complex wave vector consisting of a real part, describing the propagation in space, and an imaginary part corresponding to the material absorption. Decomposition gives

$$\tilde{k}(\omega) = k_0 + \Delta k(\omega) + i \frac{\alpha(\omega)}{2}$$

with  $k_0$  the constant wave vector given by the classic refractive index,  $\Delta k(\omega)$  describing phase-dependent effects such as dispersion, and  $\alpha(\omega)$  the absorption coefficient. If now the spectral amplitude of a measurement is divided by that of a reference, the amplitude transmission containing the amplitude and phase information can be derived:

$$t = \frac{\tilde{E}(\omega)_{\text{sample}}}{\tilde{E}(\omega)_{\text{reference}}} = e^{-\alpha(\omega)l/2} e^{i[k_0 + \Delta k(\omega)]l}$$

In an intensity transmission measurement

$$T = |t|^2 = \frac{I(\omega)_{\text{sample}}}{I(\omega)_{\text{reference}}} = e^{-\alpha(\omega)l}$$

the phase information is lost.

This procedure is done to obtain spectroscopic information on the sample under test. Additionally, the measurements of the electric field in the time domain also open time-of-flight applications like thickness measurements. These measurements can be done in reflection and transmission.

By adding a sample into the terahertz path, time delay is induced. Let  $t_{sample}$  and  $t_{reference}$  be, respectively, the time delay of the sample and reference pulse, for example, the temporal position of the peak maximum. Then the time delay between these two pulses  $\Delta t$  is proportional to the sample thickness  $D$ :

$$\Delta t = t_{sample} - t_{reference} = \frac{(n - 1)}{c} D$$

with  $c$  the vacuum velocity of light and  $n$  the refractive index of the sample in the terahertz range. Typical refractive indices, for example, are 1.5 for common polymers. The refractive index  $n$  is either determined by measuring a sample of known thickness or by checking literature values. So through the evaluation of the pulse delay  $\Delta t$  caused by a sample with known refractive index  $n$ , the thickness  $D$  can be easily calculated. Depending on parameters such as absorption of the sample and temporal resolution, layers can be resolved down to a thickness of a few micrometers in transmission mode.

A measurement geometry working in reflection mode is also possible. At each interface between two media with any step of the refractive index  $n_1$  and  $n_2$ , a fraction of the terahertz amplitude  $r$  is reflected at normal incidence according to Fresnel's law:

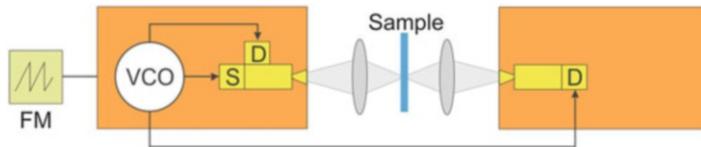
$$r = \frac{(n_1 - n_2)}{(n_1 + n_2)}$$

So at each interface with a refractive index step, a reflection occurs. By evaluation of the delay (time of flight), a depth profile of the sample can be recorded. As measurements in reflection mode can reveal the multilayer structure, the exact order of the layers in a multilayer system cannot be revealed in transmission measurement.

## Terahertz FMCW System

The frequency sweep is realized by the combination of the function generator, which supplies a sawtooth ramp, and a voltage-controlled oscillator (VCO) which drives a microelectronic multiplier chain for terahertz generation and a single-diode-based harmonic mixing module. Within the mixer, the VCO signal is multiplied and superimposed with the received terahertz signal for heterodyne detection.

As depicted in Fig. 2, the terahertz radiation is sent and received through a single horn antenna by using a directional coupler, which is connected to the source and the detector unit. Additionally, the signal of the VCO can be fed to the corresponding receiver head for transmission mode measurements.

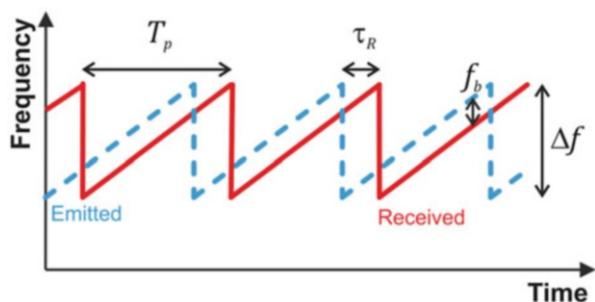


**Fig. 2** Schematic of the FMCW setup in focused configuration; a function generator (FG) is used to linearly ramp a voltage-controlled oscillator (VCO) which drives the transceiver (source (S) and detectors (D), respectively)

**Table 1** Typical specifications of a FMCW terahertz system

Sensor	100 GHz	150 GHz	300 GHz
<b>Frequency range</b>	70–110 GHz	110–170 GHz	320–320 GHz
<b>Dynamic range (acquisition time per pixel: 100 ms)</b>	>70 dB	>70 dB	>70 dB
<b>Spatial resolution (focal length of lens: 100 mm)</b>	3 mm	2 mm	1 mm
<b>Depth resolution in air</b>	9 mm	6 mm	3 mm

**Fig. 3** Frequency evolution of the transmitted (dashed) and received signal (solid) from one scatterer after a time delay of  $\tau_R$



The signal frequency of the VCO is modulated in a range between 13 and 18 GHz by a sawtooth ramp generator and is up-converted by the frequency multiplier chain to the operation frequency range, as detailed in Table 1. The system has been configured to achieve a single 250  $\mu$ s frequency sweep. In the conventional focused configuration, a set of quasi-optical Teflon lenses is used to focus the emitted terahertz radiation into the test sample and the transmitted and/or reflected parts of the terahertz beam into the transceiver. For the given setup, the minimum spatial resolution for each sensor pair is also listed in Table 1. The FMCW technique can be described by a sawtooth-like frequency modulation as shown in Fig. 3.

The linear frequency sweep of the emitted terahertz radiation allows the acquisition of the time delay  $\tau_R$  between the transmitted (TX) and received (RX) signal by superposition. This results in a beat signal with a frequency  $f_b$  equal to the difference frequency of the superimposed signals, which is proportional to the optical path difference in a medium with refractive index  $n$ . In respect to the period  $T_p$  and the

bandwidth  $\Delta f$  of the frequency modulation, the distance or depth of the inspected region is described by

$$d = \frac{c}{2n} \tau_R = \frac{cf_b}{2n\Delta f}$$

The achievable depth resolution is given by the bandwidth of the frequency modulation. The use of the FMCW technique on terahertz transparent materials therefore allows the reconstruction of tomographic terahertz images. Commonly, the signal path differences between the received terahertz radiation and the reference signal from the VCO generate beat signals with constant frequencies at the receiver output, while the abovementioned calibration procedure ensures the required quasilinear frequency sweep of the VCO. The received mix of beat signals from different depths can be separated by Fourier analysis. The resulting spectrum reveals the amplitude of signals retrieved from different depths. For the conventional focused system configuration, it has to be taken into account that the achievable image resolution varies with the inspected depth as the beam radius increases with the distance to the beam waist. Hence, the stronger optical signal and better lateral resolution are preserved with a tighter focus, but with a smaller depth of focus (Rayleigh length), and therefore a reduced range in which high-resolution images can be acquired.

## Imaging Concepts

In many applications, it is required to generate an image of the object under test. The advantage of using terahertz waves is that many electrically nonconductive materials like ceramics and polymers are transparent in this spectral region. Therefore, not just an image of the surface is generated, but also a view inside or even though the object is possible. Another advantage is that a phase-sensitive detection is used in most of the terahertz techniques like TDS or FMCW, and in this way three-dimensional information on samples is obtained. In contrast to ultrasonics, there is no coupling media, and to X-ray the health risk due to ionizing radiation is eliminated.

Passive terahertz systems detect the natural radiation emission of objects due to their temperature and chemical composition (Fetterman et al. 2007; Luukanen et al. 2007; Martin et al. 2007; May et al. 2013). Because the power of this radiation is extremely low, the detectors are very sensitive and mostly cooled down to 4 K and below. These detectors do not measure phase, but only the total intensity of the absorbed radiation. Due to the operation conditions, they are used in astronomy and security, but not in NDT applications. These detectors are not considered in the following discussion.

Active systems have an artificial source of terahertz radiation, which is used to illuminate the objects. The detection scheme can be either phase- or power-sensitive. In the latter case, no information on time of flight is available, and the approach is used in transmission measurement only. With phase-sensitive measurements, time-of-flight information can be used to reveal the spatial structure of the object. This approach is applied in reflection and transmission setups (Friederich et al. 2011).

Most of the terahertz detectors are single-element sensors. To obtain an image using such a sensor either the detector (and often also the source) or the object has to be moved with respect to the other (Fig. 4). These raster scans are very precise, but time-consuming. In some cases, the speed of moving the detector or the samples is the limiting time factor for the scanning. One approach to overcome this speed issue is the use of a combination of rotating polygon scanner and f-theta optics (see Fig. 4, right). While the systems measure along a line, the sample moves with moderate speed. The disadvantage of this setup is the large optics, which has to be used for these long wavelengths.

A few multi-pixel cameras are available, either commercially or in research laboratories. These cameras have a very limited field of view and detect only power; no reflection measurements are possible. Depending on the detection scheme, the sensitivity of the different camera types varies strongly. Due to the limited field of view, a raster scanning is often necessary to generate a complete image of an object. As only the power of the transmitted terahertz waves is detected, no structural information like depth or thickness is available. Therefore, the cameras are often used in niche applications like beam profiling.

Depth information and reflection measurements are often required in industrial applications. Therefore, phase-sensitive detectors have to be applied. A two-dimensional camera with a sufficiently large field of view requires a large number of pixels or sensors. As the actual cost per pixel is rather high and the size of such an array is quite large due to the imaging conditions, these cameras have not been realized. To reduce the costs, another approach is more feasible: a combination of a linear array and linear movement of the sample (see Fig. 5).

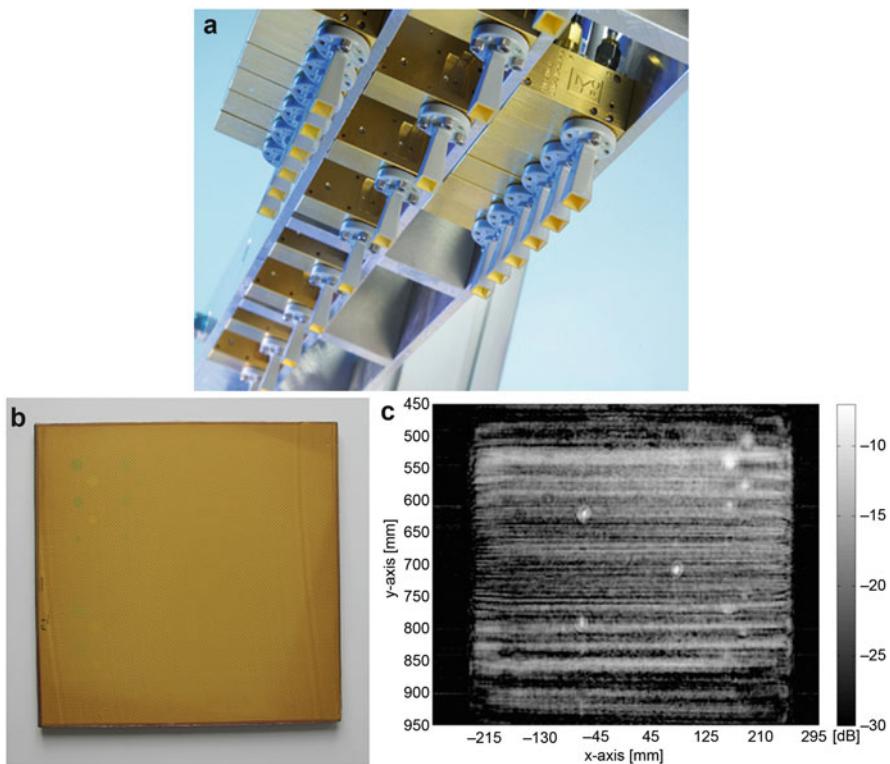
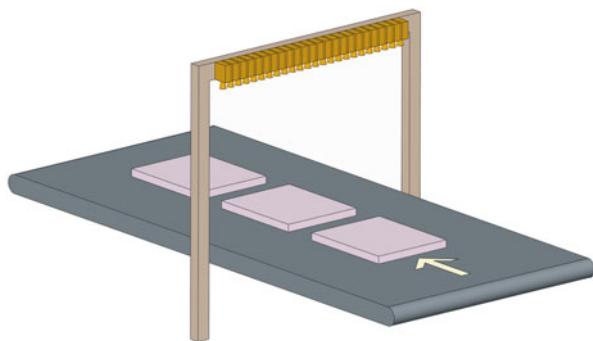
The advantage of the abovementioned imaging approaches is that images are automatically sharp as in photography. Still the linear camera requires a large number of sensors. New concepts to overcome this drawback are in development. One of the most promising approaches is the MIMO method (MIMO: multiple-input multiple-output). In this approach, the hardware costs are greatly reduced, and the depth of focus is significantly improved through the use of sparse arrays and a synthetic focus. Simplifying, the basic idea is that one source emits terahertz waves and all detectors are phase-sensitive to the receiving signal. This procedure is repeated until all sources have been used. Afterward, the image has been



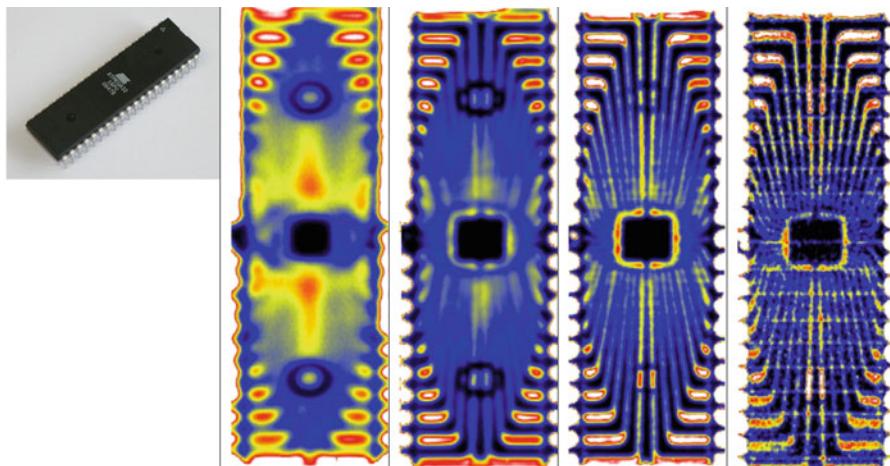
**Fig. 4** Schemes of raster scans. Left: XY scanner using linear axes. Right: scanner using polygon mirror and f-theta optics

reconstructed using the information of the geometry of the array. The example in Fig. 6 shows a MIMO radar system with 12 senders and 12 receivers at 60–110 GHz (Baccouche et al. 2017). This approach can be transferred easily to higher frequencies.

**Fig. 5** Scheme for imaging approach. Linear camera and conveyor belt



**Fig. 6** MIMO radar system with 12 senders and 12 receivers at 60–110 GHz. Top: photograph. Bottom: reconstruction of the panel with internal defects – (a) photograph of the sample, no defect is visible; (b) reconstructed image, internal defects are visible

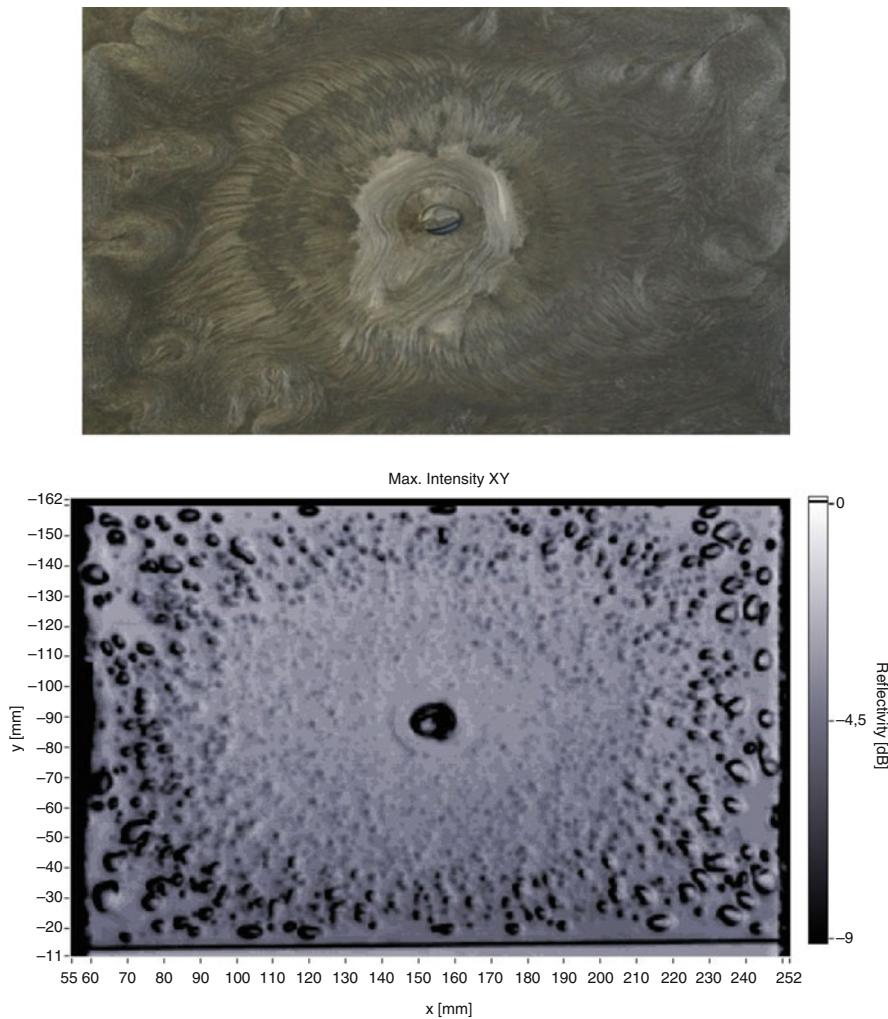


**Fig. 7** Transmission image of IC: integrated power for different spectral bands: from left to right, 0.1–0.5, 0.5–1.0, 1.0–1.5, and 1.5–2.0 THz

The broadband feature of TDS offers an additional option in hyperspectral imaging. Every pixel in the image contains the full spectral information. In transmission images, the total integrated power or the maximum amplitude of the received signal is often shown. Alternatively, the integrated power of a limited spectral band can be used for the image (see Fig. 7). The higher the frequency, the lower is the general transmission through the materials and the higher is the spatial resolution. These two effects can be seen in the images of the different spectral bands.

Based on the wavelength and the optics used to form the terahertz beam, the spatial resolution is in the range of 100  $\mu\text{m}$  to several mm. If the samples have defects or structures which are of this order of magnitude or larger, they can be detected with the correct size. If they are comparable or smaller, these structures could be detected in advantageous condition, but the size is wrong.

The features in the terahertz images are the result of scattering, fraction, absorption, and reflection. The detection of metals and other conductive materials is easily done in transmission measurement due to the total blocking of the transmitting wave. If structures (defects are also structural features) have an irregular shape, their detection is easier in transmission than in reflection mode. The reason for this is that the reflected signal is mostly not reflected directly into the detector but in arbitrary directions. In transmission, the scattering effect at the edges of structures gives sufficient contrast to reveal the structures. Figure 8 shows an example of a polymer plate, which has many air bubbles inside due to a non-optimized injection molding. In reflection mode the largest bubbles can be detected, but in transmission mode even the small air bubbles can be seen. It is noticeable that the bubbles appear as ellipsoids with a dark edge, which is caused due to the strong scattering.



**Fig. 8** Polymer plate with air bubbles inside. Top: photograph. Bottom: transmission image at 300 GHz

## Current Applications

Currently, some rather promising, industrially relevant applications are in the pipeline. All indications are that the “rollout” of terahertz applications has already begun with measurement systems. The most notable ones are nondestructive testing of coating thickness and plastic components, spectroscopy in the areas of process

control and safety monitoring, and near-field scanning of the conductivity of semiconductor design elements.

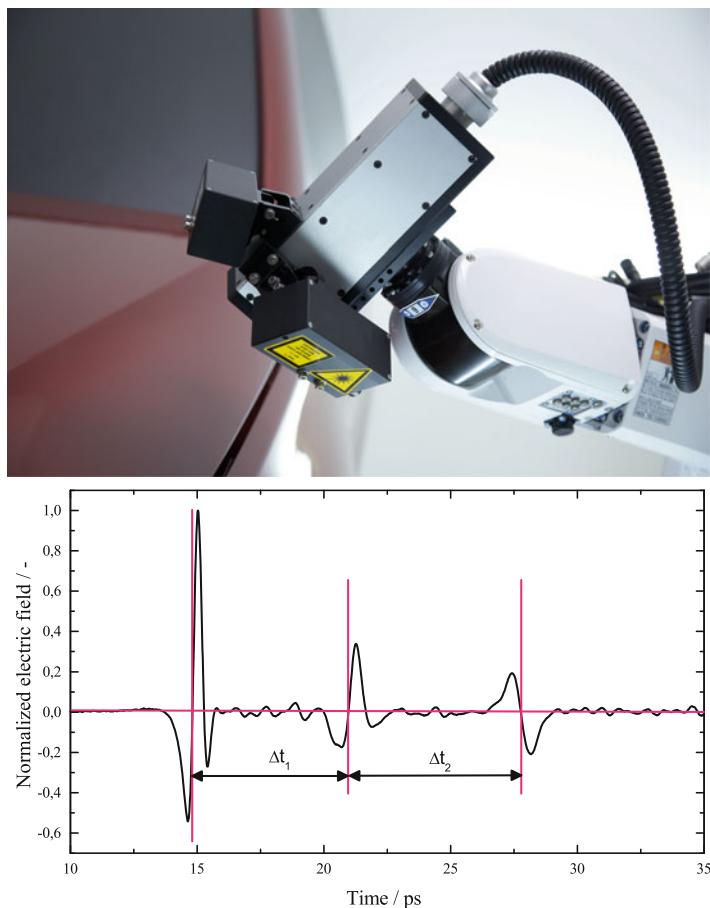
A different terahertz system has been developed for each of these applications: the pulsed optical systems occupy a wide spectrum and have very short pulses, which enables run-time analyses and spectroscopic applications. Electronic systems combine very high frequencies with a narrow spectral band and are well suited for large-volume imaging and defect detection.

The individual applications and optimal approaches are discussed in detail below.

At present, the most important application for the pulsed optical terahertz systems is for noncontact **coating thickness measurement**, needed for complex multilayer applications (Catapano et al. 2017; Dong et al. 2017; Jonuscheit 2014a; Pfeiffer et al. 2018; Tasseva et al. 2017; Yasuda et al. 2006; Yasui et al. 2005) as seen, for example, in the automobile sector (see Fig. 9). Besides the optical appearance, paint coatings serve an important function in protecting the underlying components – and not only in the automobile sector. Every separate layer has a different function, where thickness plays a critical role in achieving the desired effect. Consequently, the thickness of each layer within a multilayered system must be measured, if possible, in a noncontact and nondestructive manner. The terahertz analysis can reliably determine thicknesses even in a four-layered system – with similar material parameters in each of the layers. Depending on the system used, measurement errors are less than 1 μm. The method is considered a breakthrough as it calibrates directly on the measured coating system and requires no prior calibration on specially prepared samples. This results in reduced setup times and cost savings. Thanks to the use of its noncontact function and measuring times of less than 1 s, such systems can be used directly on the production line for 100% control.

The coatings applied on wind turbines, especially for offshore operations, shield the underlying composite material components from the harsh climatic influences. Reliable control of preassembly coatings and initial operation is absolutely essential to ensure proper operation without endangering the units. The terahertz measurement of coating thickness allows reliable, nondestructive checks of the ceramic coatings on the turbine blades, which protect the underlying metal from overheating. This capability is relevant for power plants as well as in aircraft engines. The correct thickness can now be determined without the usual destructive micrographic imaging, and immediate decisions, in the case of a layer being too thin, can be taken to apply another coat or, if too thick, to remove and reapply a new coat. Compact terahertz sensors can determine layer thickness, and terahertz technologies guarantee 100% control even with relatively complicated geometries.

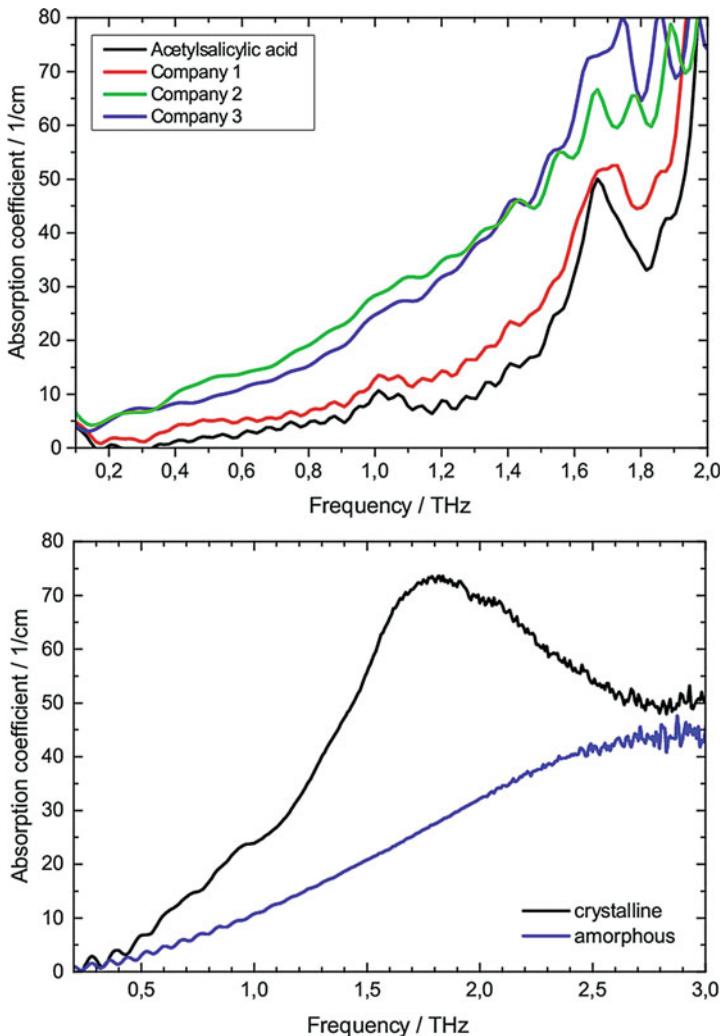
Terahertz systems have truly established themselves for use in **security systems**, one of the first areas proposed for terahertz research. Many substances like drugs or explosives exhibit characteristic absorption bands, which makes it possible to clearly identify them in the terahertz frequency bands (Ortolani et al. 2008; Shen et al. 2005). Of course, in principle this is also possible in the lab with Fourier transformation infrared spectroscopy or Raman analyses, but the terahertz systems stand out from the crowd of established technologies. The reason is that suspicious substances are often sent in letters or small packages that are inaccessible to these established technologies.



**Fig. 9** Terahertz technology is suitable in many different applications for layer thickness measurements. This approach primarily has the greatest advantages in the analysis of complex multilayer systems, for example, those found in lightweight design materials, ceramics, coatings, and films. The measurement can be performed manually with handheld devices as well as with robot assistance (top). The principle of the thickness measurement is a time-of-flight analysis (bottom)

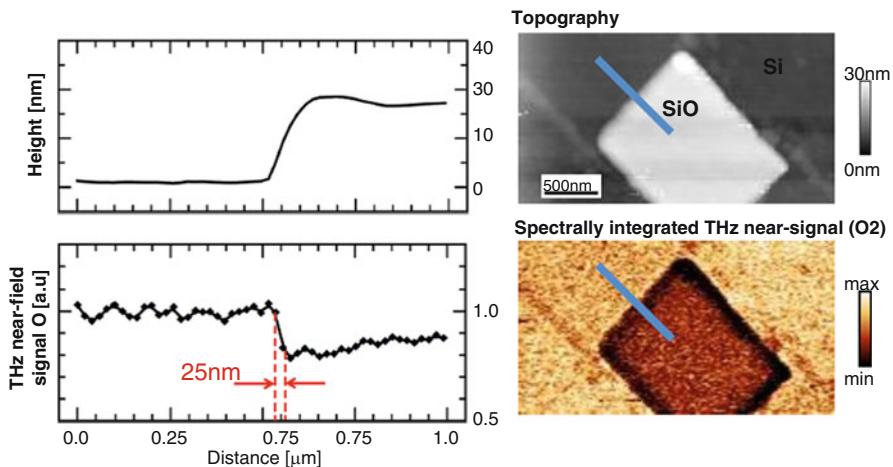
Paper and cardboard are transparent to terahertz radiation, so it is possible to check sealed items for suspected materials. The first products are already on the market (Hübner Photonics). In addition, terahertz spectroscopy is used for structural analysis in the pharmaceutical and chemical industries (Consolino et al. 2017; Reid et al. 2010; Sibik and Zeitler 2016; Wallace et al. 2008) (see Fig. 10). Compared to Raman and IR spectroscopy, terahertz systems provide a particularly high sensitivity to structural changes. Among other things, this technology permits the differentiation of various hydrate states and isomers (Jonuscheit 2014b; Theuer et al. 2011).

The field of **materials research** is presently redefining itself. The terahertz spectral range was regarded for many years as unsuitable for the analysis of small-scale



**Fig. 10** Terahertz spectroscopy enables a chemical analysis of pharmaceuticals – for example, through a sealed blister pack made of paper or plastic. Top: products from different supplier of the same active ingredient; bottom: same active ingredient in crystalline and amorphous state

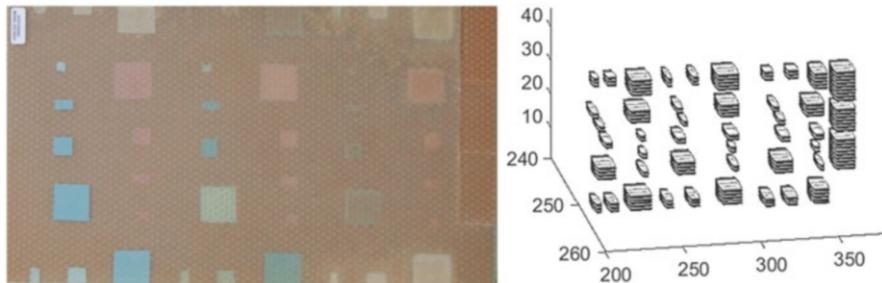
material distributions because the achievable spatial resolution in the far field is determined by the wavelength; the local resolution was typically  $100 \mu\text{m}$ . However, this boundary has now been overcome in combination with the use of a scattered near-field microscope. Full spectral resolution in the terahertz range enables material-sensitive imaging that has demonstrated spatial resolution better than  $30 \text{ nm}$ , effectively less than 1000th of the wavelengths used (Fig. 11). Such high-end devices open up completely new perspectives for scientists in materials research and are also commercially available (Böhmger et al. 2016; Liewald et al. 2018).



**Fig. 11** Microscopic scanning images and terahertz near-field measurements on semiconductor test structures. At full spectral resolution in the terahertz range, material-sensitive imaging with spatial resolutions of better than 30 nm can be achieved. (Source: Neaspec GmbH/Fraunhofer ITWM)

Modern lightweight designs increasingly use complex composite materials to achieve a balance between low weight and high strength and stiffness. Possible composites for use in aircraft construction are mainly sandwich designs, consisting of two or more thin glass fiber-reinforced plastic plates with intermediate layers of foam or honeycomb structures. The **nondestructive testing of lightweight designs** represents a nearly impossible challenge for the established testing methods, but the application of terahertz technique has been demonstrated successfully here (Chen et al. 2017; Cristofani et al. 2014; Dong et al. 2015; Friederich et al. 2018; Kaimal et al. 2018; Ospald et al. 2014; Stoik et al. 2008). During the European Union-sponsored DOTNAC project, the terahertz technology was in direct competition with the established methods. Samples were prepared with defects and tested using different techniques. The terahertz technology proved itself to be an alternative to be given serious consideration: terahertz systems were able to detect, localize, and identify all defects (see Fig. 12). Electronic waveguide systems were used in the project because good depth resolution is necessary to localize the defects (Cristofani et al. 2014).

Most plastic products are manufactured by melting granulate and subsequently shaping them by means of injection molding, extrusion, blow molding, or simple casting. In the example presented here, polymer is poured onto a rolling mill after melting. To enhance the mechanical properties, threads are spread around and along the length of the roller before becoming engulfed entirely by the liquid polymer. After casting and cooling, the surface is machined, and, at the end of the production process, the processed polymer tube is removed from the roller. One quality indicator of the polymer tube is the presence (or absence) of bubbles. During the casting process, bubbles may form near the surface which can be detected from the outside, thus allowing a decision to immediately halt the production. However, internal



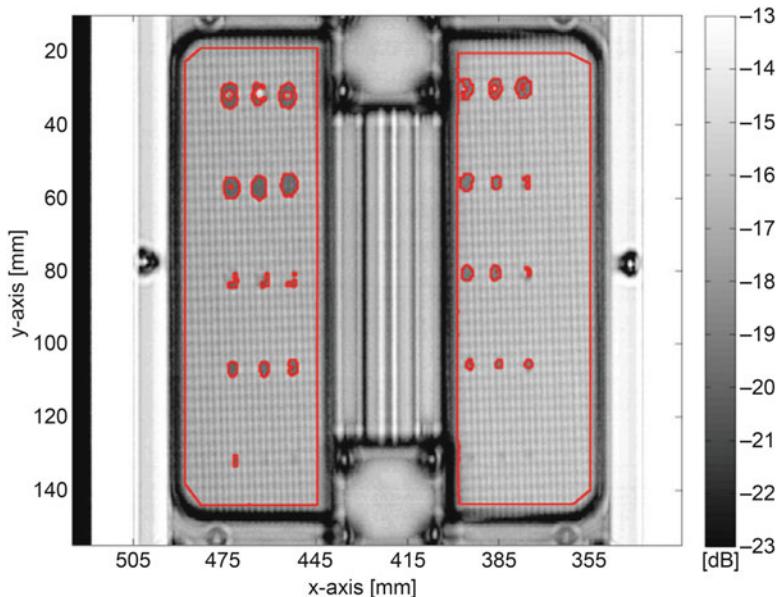
**Fig. 12** A sandwich reference sample with honeycomb structure. Left: photo of the sample. Right: a 3-D display of detected defects for measurements at 150 GHz

bubbles are only visible at the end of a complete production process. The late control check means that in the event of defects, many resources like material and machine time have already been consumed and the start of the replacement production is delayed. This provides the grounds to search for an **inline inspection system** that can already be used during the casting process. Electronic terahertz systems hold the promise of noncontact measurements (shortly after casting while the plastic is still soft) and fast measurement times (appropriate to the production speed).

Flaws were introduced on the far side of the sample (the side directly on the roller in the case of the original) to determine the smallest yet to be detected defects, and then the sample was stretched onto a metal plate so the defects are in contact with the plate. Two rows each with three holes drilled to the same size from 5 mm to smaller than 1 mm were produced; one row of holes is drilled with a flat bottom drill bit, and the second row is drilled with a pointed drill bit. The smooth shape of the flat bottom holes is easier to detect than the pointed holes, which have less direct reflection. The holes are drilled to a depth of less than 1 mm. All of the holes can be detected when the image is evaluated by a trained analyst; and, with some minor adjustment to the image analysis software (OpenCV), nearly all holes can be detected automatically, as shown in Fig. 13. The red markings show structures/faults detected by the software. In the lower right of the image, three less distinct areas can be seen that represent the smallest holes drilled with the sharp bit. The potential of automatic image analysis is clearly demonstrated.

## Outlook

All of the terahertz applications mentioned here are from the area of measuring systems. This is where the terahertz technology first came into its own and achieved industrial market maturity. Other areas will soon follow, in particular, the information and communication sector (Nagatsuma et al. 2016). The trend is clear and uninterrupted: terahertz systems are becoming increasingly capable and more efficient. The terahertz technology is gaining further momentum in connection with improved algorithms for data evaluation. The potential of terahertz inspection



**Fig. 13** Automated defect detection results. A defect has been detected in the areas shown in red. The reinforcing threads are clearly seen as horizontal and vertical lines

technology to expand into additional market sectors is huge and ensured by the development of application-oriented systems. In summary, terahertz technology is clearly entering a period of very promising activity for the future.

## References

- Baccouche B, Agostini P, Mohammadzadeh S, Kahl M, Weisenstein C, Jonuscheit J, Keil A, Löffler T, Sauer-Greff W, Urbansky R, Bolivar PH, Friederich F (2017) Three-dimensional terahertz imaging with sparse multistatic line arrays. *IEEE J Sel Top Quantum Electron* 23:8501411
- Böhmler M, Huber A, Eisele M (2016) THz nano-spectroscopy with 25 nm spatial and 10 f. time resolution. In: The 41st international conference on infrared, millimeter, and terahertz waves (IRMMW-THz) 16502507
- Bründermann E, Hübers H-W, Kimmitt MFG (eds) (2012) Terahertz techniques. Springer, Berlin/Heidelberg
- Catapano I, Soldovieri F, Mazzola L, Toscano C (2017) THz imaging as a method to detect defects of aeronautical coating. *J Infrared Millimeter Terahertz Waves* 3810:1264–1277
- Chen J, Wang J, Cui H (2017) Nondestructive evaluation of glass fiber honeycomb sandwich panel using reflective terahertz imaging. *J Sandw Struct Mater* 0:1–13. <https://doi.org/10.1177/1099636217711628>
- Consolini L, Bartalini S, de Natale P (2017) Terahertz frequency metrology for spectroscopic applications: a review. *J Infrared Millimeter Terahertz Waves* 3811:1289–1315

- Cristofani E, Friederich F, Wohnsiedler S, Beigang R (2014) Non-destructive testing potential evaluation of a THz frequency-modulated continuous-wave imager for composite materials inspection. *Opt Eng* 53(03). <https://doi.org/10.1117/1.OE.53.3.031211>
- Dietz RJB, Vieweg N, Puppe T, Zach A, Globisch B, Göbel T, Leisching T, Schell M (2014) All fiber-coupled THz-TDS system with kHz measurement rate based on electronically controlled optical sampling. *Opt Lett* 39(22):6482–6485
- Dong J, Kim B, Locquet A, McKeon P, Declercq N, Citrin DS (2015) Nondestructive evaluation of forced delamination in glass fiber-reinforced composites by terahertz and ultrasonic waves. *Compos Part B Eng* 79:667–675
- Dong J, Wu X, Locquet A, Citrin DS (2017) Terahertz super-resolution stratigraphic characterization of multi-layered structures using spares deconvolution. *IEEE Trans Terahertz Sci Technol* 7:260–267
- Fetterman MR, Grata JA, Dinu R, Koenig M, Visnansky AD, Kiser WL (2007) Electro-optic polymer modulators as passive mm wave detectors. *Proc SPIE* 6472. <https://doi.org/10.1117/12.701042>
- Friederich F, von Spiegel W, Bauer M, Meng F, Thomson MD, Boppel S, Lisauskas A, Hils B, Krozer V, Keil A, Löffler T, Henneberger R, Huhn AK, Spickermann G, Bolivar PH, Roskos HG (2011) THz active imaging systems with real-time capabilities. *IEEE Trans Terahertz Sci Technol* 1:183–200
- Friederich F, May KH, Baccouche B, Mattheis C, Bauer M, Jonuscheit J, Moor M, Denman D, Bramble J, Savage N (2018) Terahertz radome inspection. *Photonics*. <https://doi.org/10.3390/photonics5010001>
- HÜBNER Photonics. [http://www.hubner-terahertz.de/Produkte/T\\_COGNITION.html](http://www.hubner-terahertz.de/Produkte/T_COGNITION.html). Accessed 29 May 2018
- Jaeschke T, Bredendiek C, Pohl N (2013) A 240 GHz ultra-wideband FMCW radar system with on-chip antennas for high resolution radar imaging. *IEEE MTT-S Int Microw Symp*. <https://doi.org/10.1109/MWSYM.2013.6697495>
- Jonuscheit J (2014a) Zerstörungsfreie Analyse – Schichtdicken von Mehrsichtsystemen online messen. *QZ Jahrg* 59:94–96
- Jonuscheit J (2014b) Strukturanalyse mittels Terahertz. *GIT Labor Fachz* 5:27–29
- Kaimal H, Devi N, Ray S, Rajagopal P, Balasubramanian K, Pesala B (2018) Non-destructive evaluation of GFRP-wood sandwich structure composite using terahertz imaging. *Proc SPIE* 10531. <https://doi.org/10.1117/12.2289718>
- Klatt G, Nagel M, Dekory T, Bartels A (2009) Rapid and precise read-out of terahertz sensors by high-speed asynchronous optical sampling. *Electron Lett* 45:310–311
- Kolano M, Gräf B, Weber S, Molter D, van Freymann G (2018) Single-laser polarization-controlled optical sampling system for THz-TDS. *Opt Lett* 43(6):1351–1354
- Liewald C, Mastel S, Hesler J, Huber AJ, Hillenbrand R, Keilmann F (2018) All-electronic terahertz nanoscopy. *Optica* 5(2):159–163
- Lisauskas A, Pfeiffer U, Öjefors E, Bolivar PH, Glaab D, Roskos HG (2009) Rational design of high-responsivity detectors of terahertz radiation based on distributed self-mixing in silicon field-effect transistors. *J Appl Phys* 105. <https://doi.org/10.1063/1.3140611>
- Luukanen A, Grönberg L, Helistö P, Penttilä JS, Seppä H, Sipola H, Dietlein CR, Grossman EN (2007) Passive Euro-American terahertz camera (PEAT-CAM): passive indoors THz imaging at video rates for security applications. *Proc SPIE* 6548. <https://doi.org/10.1117/12.719778>
- Martin CA, Lovberg JA, Dean WH, Ibrahim E (2007) High resolution passive millimeter-wave security screening using few amplifiers. *Proc SPIE* 6548. <https://doi.org/10.1117/12.718950>
- May T, Heinz E, Peiselt K, Zieger G, Born D, Zakosarenko V, Brömel A, Anders S, Meyer H-G (2013) Next generation of a sub-millimetre wave security camera utilising superconducting detectors. *IOP Publ J Instrum* 8. <https://doi.org/10.1088/1748-0221/8/01/P01014>
- Mittleman D (ed) (2003) Sensing with terahertz radiation. Springer, Berlin/Heidelberg
- Naftaly M (ed) (2015) Terahertz metrology. Artech House Publishers, Boston/London

- Nagatsuma T, Ducournau G, Renaud CC (2016) Advances in terahertz communications accelerated by photonics. *Nat Photonics* 10:371–379. <https://doi.org/10.1038/nphoton.2016.65>
- Öjefors E, Lissauskas A, Glaab D, Roskos HG, Pfeiffer UR (2009) Terahertz imaging detectors in CMOS technology. *J Infrared Millimeter Terahertz Waves* 30:1269–1280. <https://doi.org/10.1007/s10762-009-9569-4>
- Ortolani M, Lee JS, Schade U, Hübers H-W (2008) Surface roughness effects on the terahertz reflectance of pure explosive materials. *Appl Phys Lett* 93:081906
- Ospald F, Zouathi W, Beigang R, Matheis C, Jonascheit J, Recur B, Guillet J-P, Mounaix P, Vleugels W, Bosom PV (2014) Aeronautics composite material inspection with a terahertz time-domain spectroscopy system. *Opt Eng SPIE* 533:031208
- Pfeiffer T, Weber S, Klier J, Bachtler S, Molter D, Jonascheit J, von Freymann G (2018) Terahertz thickness determination with interferometric vibration correction for industrial applications. *Opt Express* 2610:12558–12568. <https://doi.org/10.1364/OE.26.012558>
- Reid CB, Pickwell-MacPherson E, Laufer JG, Gibson AP, Hebden JC, Wallace VP (2010) Accuracy and resolution of THz reflection spectroscopy for medical imaging. *IOP Publ* 5516:4825–4838
- Shen YC, Lo T, Taday PF, Cole BF, Tribe WR, Kemp MC (2005) Detection and identification of explosives using terahertz pulsed spectroscopic imaging. *Appl Phys Lett* 86:241116. <https://doi.org/10.1063/1.1946192>
- Sibik J, Zeitler JA (2016) Direct measurement of molecular mobility and crystallization of amorphous pharmaceuticals using terahertz spectroscopy. *Adv Drug Deliv Rev* 100:147–157
- Skolnik M (2008) Radar handbook, 3rd edn. McGraw-Hill Education, New York
- Stoik CD, Bohn MJ, Blackshire JL (2008) Nondestructive evaluation of aircraft composites using transmissive terahertz time domain spectroscopy. *Opt Express* 162:17039–17051
- Tasseva J, Bartolini P, Tascin A, Striova J (2017) Thin layered drawing media probed by THz time-domain spectroscopy. *Analyst* 142:42–47. <https://doi.org/10.1039/C6AN02113A>
- Tessmann A, Leuther A, Massler H, Hurm V, Kuri M, Zink M, Riessle M, Stulz HP, Schlechtweg M, Ambacher O (2014) A 600 GHz low-noise amplifier module. *Int Microw Symp.* <https://doi.org/10.1109/MWSYM.2014.6848456>
- Theuer M, Harsha SS, Molter D, Torosyan G, Beigang R (2011) Terahertz time-domain spectroscopy of gases, liquids, and solids. *ChemPhysChem* 12:2695–2705. <https://doi.org/10.1002/cphc.20110158>
- Wallace VP, MacPherson E, Zeitler JA, Reid C (2008) Three-dimensional imaging of optically opaque materials using nonionizing terahertz radiation. *J Opt Soc Am A* 2512:3120–3133
- Wilk R, Hochrein T, Koch M, Mei M, Holzwarth R (2011) OSCAT: novel technique for time-resolved experiments without moveable optical delay lines. *J Infrared Millimeter Terahertz Waves* 325:596–602
- Yasuda Y, Yasui T, Araki T, Abraham E (2006) Real-time two-dimensional terahertz tomography of moving objects. *Opt Commun* 267:128–126
- Yasui T, Yasuda T, Sawanaka K, Araki T (2005) Terahertz paintmeter for noncontact monitoring of thickness and drying progress in paint film. *Appl Opt* 4432:6849–6856



# Ground Penetrating Radar

28

X. Lucas Travassos and Mario Fernandez Pantoja

## Contents

Introduction .....	988
Overview of a GPR System .....	989
Fundamental Theory of GPR .....	991
Electromagnetic Wave Propagation .....	992
Material Properties .....	994
Antennas .....	996
System Specification .....	999
A practical application: GPR detection in concrete slabs .....	1004
Postprocessing Support Tools (PSTs) .....	1010
Signal and Image Processing Techniques .....	1013
Pattern Recognition .....	1016
Summary .....	1020
References .....	1021

## Abstract

In recent years, the nondestructive testing (NDT) of structures and soils using ground penetrating radars (GPRs) has become a mature technology. The particular interest in this technique is explained by several advantages when compared to other NDT techniques: the portability of the equipment because of its moderate weight, relative low cost of the survey, reasonable budget of the initial investment, and high versatility in terms of application for different purposes and scenarios. However, the success of GPR surveys is not straightforward due to the complexity of the physical phenomena involved. The determination of the

---

X. Lucas Travassos (✉)

Federal University of Santa Catarina, Joinville/Florianopolis, Brazil  
e-mail: [lucas.travassos@ufsc.br](mailto:lucas.travassos@ufsc.br)

M. F. Pantoja  
Universidad de Granada, Granada, Spain  
e-mail: [mario@ugr.es](mailto:mario@ugr.es)

most appropriate hardware configuration (mostly, antennas and particular waveforms of electromagnetic pulses) is determined by a thorough analysis of the application (e.g., material properties and features of the buried object and host medium). In this way, an initial choice among the available electronic equipment is made by gathering the relevant information of the survey: the expected depth of the target embedded, the resolution required to identify targets of a definite size, the physical contrast between constitutive parameters of flaw and host medium, and the signal to noise ratio of the measurement due not only to electronics but also to inherent conditions of the survey (i.e., existing clutters, nonhomogeneities of the host medium, and roughness of the interface air-host). Therefore, a GPR survey is considered a multidisciplinary problem requiring contributions: from electrical engineering, to design and manufacture versatile and powerful equipment; from physics, to analyze and choose an optimum configuration for the desired application; and from computer science, to achieve proper predictions through the information provided by the measurements. This chapter introduces briefly this multidisciplinary approach by presenting first the electromagnetic phenomena leading to the detection with GPR, then by listing the characteristics of available equipment, and finally by enumerating some of the latest computer techniques for the discovery and classification of targets.

---

## Introduction

The idea of mapping buried objects (targets) in the soil (host medium) through scattering of electromagnetic waves was first envisioned by Hulsenbeck in 1926 (Hulsenbeck et al. 1926). Further efforts from scientists working on remote sensing led to a consistent theoretical formulation of the problem which revealed the extraordinary potential of GPR as a detection technique (Cook 1975; Evans 1963). At that early stage, all basic requirements for an efficient operation of GPR were identified: the electromagnetic energy has to be efficiently transmitted through the host medium, the electromagnetic pulse has to penetrate sufficiently into the host medium, the electromagnetic contrast between host and target medium has to be large enough to provide a clear scattered pulse, and the bandwidth of this pulse has to be large enough to provide resolution according to the size of the target. In this sense, GPRs share goals and challenges with most of the usual NDT methods, this is, the shape and size of any target is intended to be detected at any depth in any host medium including those with characteristics similar to those of the target.

However, the GPR technique becomes widely employed, being considered nowadays as one of the most powerful NDT tools, even when compared to seismic, acoustic, infrared, and optical sensors. The unique advantages offered by GPR are related to the overall performance achieved by electromagnetic microwaves: with reasonable power requirements, they combine a reasonably large penetrating range (more than optical and infrared sensors but less acoustic and seismic ones) with sufficient spatial resolution (greater than acoustic and seismic but lower than infrared and optical techniques) and a fair contrast between host and target media. But the

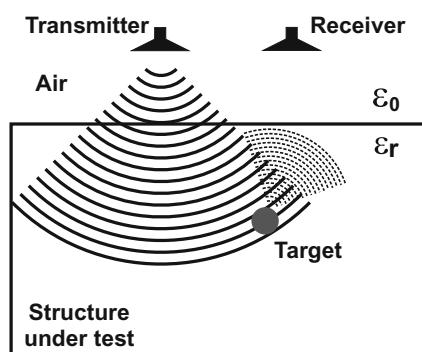
ability to detect a broad range of targets also presents drawbacks, because inhomogeneous soils or additional scattering nontarget objects (known as clutters) interact with the scattered field of the actual target, masking thus the detecting pulse. To avoid this issue scanning in two or three dimensions at different frequencies is commonly made, leading to collection of a high number of measurements which require specialized postprocessing techniques to complete the survey.

The aim of this chapter is to describe the most relevant GPR features, the electromagnetic wave propagation and scattering phenomena involved, and some of the latest techniques to improve the mapping of buried objects in a host medium. The organization of the chapter is as follows. Section “[Introduction](#)” provides an overview of the GPR scenario and discusses relevant information for the design of any survey. Section “[Fundamental Theory of GPR](#)” explains the electromagnetic phenomena involved and the main characteristics of commercial GPRs. Finally, some postprocessing support tools are presented in section “[Postprocessing Support Tools \(PSTs\)](#),” aimed to improve the performance and success rate of GPR surveys.

## Overview of a GPR System

A typical GPR scenario is shown in Fig. 1. The GPR operating principle may be divided into four parts: (a) the electromagnetic waves from a transmitting antenna are directed into the host medium under test; (b) the refracted waves propagate through the host reaching a boundary of different electrical characteristic (named flaw or target); (c) a typical reflection-refraction phenomenon arises, where the scattered wave reflects partly the incident energy and the remainder energy travels through the flaw at a different velocity to greater depths in the host; and (d) this scattered wave (called echo) travels back to the receiving antenna. In summary, the physics of GPRs is based on the reflection, refraction, and scattering processes, which transform the incident wave into a received signal with different amplitude and phase from the transmitted wave. The particular waveform is mainly determined according to the electrical contrast (i.e., ratio of relative permittivity and conductivity of target and host media) as well as the particular shape of the target.

**Fig. 1** GPR problem description



Regarding the received waveform, Fig. 2 shows a typical time-domain signal detected in GPR surveys. It can be described as the superposition of several waveforms each one representing the response of some part of the GPR system depicted in Fig. 1, in addition to noise contributions coming either from small flaws or inhomogeneous compounds in the host. It is possible to decompose the received waveform  $f_r(t)$  as (Daniels 2004):

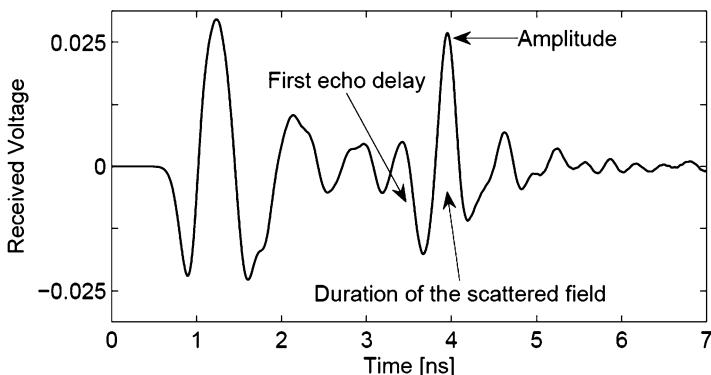
$$f_r(t) = f_a(t) + f_s(t) + sg(t) + ns(t) \quad (1)$$

where  $f_a(t)$  represents the antenna interference,  $f_s(t)$  the reflections from the ground surface,  $sg(t)$  is the signal from underground targets, and  $ns(t)$  the noise. In general,  $f_a(t) + f_s(t)$  is called the direct wave or clutter in GPR measurements.

In GPR surveys, the process of identifying the scattering objects once a set of incident and scattered waves are known (in one or, most commonly, several locations over the host medium) is called the inverse problem. In the inverse problem, the received signals are processed for detection, classification, and localization of targets.

To this end, advanced postprocessing support tools such as neural networks (Yang and Bose 2005), image processing (Donato and Crocco 2015), or wavelet transforms (Du et al. 2016) can be employed, as well as other well-known techniques such as filtering or amplification of some frequencies of interest (Daniels 2004). Additionally, received waveforms can be processed by combining spatial, frequency and/or time-domain signals, with the latest being the most common choice. An important issue, common to all these techniques, is the removal of the direct wave which is usually performed prior to application of any postprocessing strategy. In spite of the many contributions aimed at accurate predictions, the complexity of real-world scenarios (which usually involve inhomogeneous or dispersive lossy dielectrics and also irregular boundaries for the host medium) opens issues with plenty of opportunities for research.

Manufacturers of commercial GPR systems employ various approaches and techniques to assemble their systems, which can be tailored according to the application and target characteristics. In general, the main features to be considered in the variety of the GPR systems suppliers are (a) unit size and weight, (b) power



**Fig. 2** Reflected wavefield from a buried target

consumption, (c) central frequency, (d) transmitted pulse waveform (in terms of duration and modulation), and (e) number, type, and position of sensors (antennas).

The type and position of antennas according to the structure under test determine the amount and quality of measurements. According to the number and position of transmitting/receiving antennas, there are three possible operating modes: (a) monostatic, where a transmitting/receiving antenna is located in the same place over the material under test; (b) bistatic, where transmitter and receiver are at different positions; and (c) multistatic, where multiple transmitters and/or multiple receivers, usually designed to form antenna arrays, can be located in different places. Any specific choice of configuration will affect not only the amount of data but also their quality, because a higher number of antennas at closer distances will result in strong coupling between antennas, either by the direct wave or by the first reflection in the structure under test, which jeopardizes the removal of clutter at the postprocessing stage.

Additionally, the antenna type will affect greatly the performance of the GPR survey. Highly directive antennas explore narrow zones of the host at deeper distances, but, in practice, they also increase the number of scans required. Furthermore, the type of antenna limits the operational bandwidth, which has to be chosen carefully. Higher bandwidths can be achieved through ultra-wideband antennas (Balanis 2005) which avoid the problem of late-time ringing, defined as a delayed electromagnetic radiation which increases the noise in the data. On the other hand, the greater the bandwidth, the higher sensitivity to external electromagnetic interferences (e.g., broadcast antennas operating in the same frequency range as the GPR system) and the lower energy per frequency available for the scan.

In this sense, the penetration depth and resolution are also directly related to the pulse waveform and its frequency content. Time-domain or pulsed radars, commonly used in practice, are based on the radiation of a particular waveform with a high bandwidth which eases the identification of targets and flaws by the presence of scattered waves, in a process called time-windowing. On the other hand, frequency-domain GPR systems use frequency modulation where a carrier frequency scans bandwidths with a fixed step. Stepped-frequency continuous-wave systems are available in commercial GPRs, and its use is increasing due to the advantages over pulsed radars in terms of shaping the power spectral density and a higher mean power.

In summary, the choice of any GPR equipment has to be made according to the particular application and characteristics of the target and host medium. This choice also involves a trade-off between resolution and penetration depth which are the key factors for the success in the detection and classification of targets. In the following section, the electromagnetic (EM) propagation nature of GPR is briefly discussed prior to comparing the present commercial options available.

---

## Fundamental Theory of GPR

The operating principle of GPR is the propagation and scattering of electromagnetic waves in matter, offering thus an alternative prospection to other NDT methods based on energy-matter interactions of mechanical waves. Therefore, instead of

dealing with mechanical properties, GPR is based on the varying properties of electromagnetic constitutive parameters (permittivity  $\epsilon$ , conductivity  $\sigma$ , and permeability  $\mu$ ) to locate, identify, and image embedded targets in a host medium.

## Electromagnetic Wave Propagation

The nature of the electromagnetic propagation and scattering is derived from Maxwell equations which relate the electromagnetic field  $\{\mathbf{E}$ , electric field ( $V/m$ );  $\mathbf{D}$ , electric flux density ( $C/m^2$ );  $\mathbf{H}$ , magnetic field ( $A/m$ );  $\mathbf{B}$ , magnetic flux density ( $Wb/m^2$ ) $\}$ , and its sources  $\{\mathbf{J}$ , electric current density ( $A/m^2$ );  $\rho_e$ , electric charge density ( $C/m^3$ ) $\}$ . Considering a general media  $\{\epsilon, \sigma, \mu\}$ , Maxwell field equations in differential form are (Ida 2007) as follows:

Faraday's law:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (2)$$

Ampère's law:

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \quad (3)$$

Gauss' law for the electric field:

$$\nabla \cdot \mathbf{D} = \rho_e, \quad (4)$$

Gauss' law for the magnetic field:

$$\nabla \cdot \mathbf{B} = 0. \quad (5)$$

where it should be noted that on the right-hand side of (2) and (3), the time-varying electric and magnetic field also act as sources of each other. This key feature enables the propagation of electromagnetic waves through media with or without electric current or charges. Derived from these equations, some boundary conditions for the tangential and normal components of the field have to be satisfied at the interface between different media. The GPR problem can be viewed as a propagation wave problem but within the confinement of boundaries. In fact, reflection, refraction, and scattering of EM waves at boundaries provide essential means for measuring the interaction between EM waves and the structure under test, and target properties are intended to be measured through the received amplitude of electromagnetic fields at different frequencies.

Therefore, the EM wave propagation in a limited space is defined by the nature of the electric materials properties of the medium. In GPR surveys, materials can be considered as lossy and dispersive media, with constitutive parameters being scalar fields dependent on the frequency  $f$  of the wave ( $\{\epsilon = \epsilon(\omega), \sigma = \sigma(\omega), \mu = \mu(\omega)\}$ ),

where  $\omega = 2\pi f$ . In this case the relation between field intensity vectors and flux density vectors can be written as  $\mathbf{D} = \epsilon \mathbf{E}$ ,  $\mathbf{B} = \mu \mathbf{H}$ , and  $\mathbf{J} = \sigma \mathbf{E}$ . Considering these relationships in Maxwell equations, a wave equation for the electric field intensity  $\mathbf{E}$  and the magnetic field intensity  $\mathbf{H}$  can be obtained. For time-harmonic sources, the wave equation in source-free, linear, isotropic, homogeneous materials is:

$$\nabla^2 \mathbf{E} = j\omega\mu(\sigma\mathbf{E} + j\omega\epsilon\mathbf{E}) \quad (6)$$

where  $\omega = 2\pi f$  with  $f$  the frequency of the wave. For lossless nonmagnetic media, which is the usual case in GPR, the electromagnetic wave propagates at a velocity  $v$  (Ida 2007):

$$v = \frac{c}{\sqrt{\epsilon_r}} \quad (7)$$

where  $\epsilon_r$  is the relative dielectric permittivity of the medium and  $c = 3 \cdot 10^8 \text{ m/s}$  is the speed of light. As it will be further explained, the propagation wavelength  $\lambda = \frac{v}{f}$  plays a relevant role in determining the resolution capabilities of GPR.

In practice, most of the surveys contain lossy materials, which can be either conductive losses (i.e., media with a non-negligible conductivity  $\sigma$ ) or polarization losses (i.e., media with a complex permittivity  $\epsilon = \epsilon' - j\epsilon''$ ). A particularly successful case for GPR is to detect targets composed of highly conductive materials (for instance, surveys of concrete structures that contain steel reinforcement). On the other hand, when the host medium is lossy, the wave propagates with an exponentially decaying amplitude. As a consequence, the depth range of the survey can be seriously affected, and a careful choice of frequency range and transmitted power has to be made in the GPR system specification. At this point, we characterize the attenuation in lossy materials with the attenuation constant  $\alpha$ :

$$\alpha = \omega \left[ \frac{\mu\epsilon'}{2} \left( \sqrt{1 + \left( \frac{\epsilon''}{\epsilon'} \right)^2} - 1 \right) \right]^{\frac{1}{2}} \quad (8)$$

measured in nepers/meter [Np/m], and the velocity of propagation will be determined through the phase constant  $\beta$ :

$$\beta = \omega \left[ \frac{\mu\epsilon'}{2} \left( \sqrt{1 + \left( \frac{\epsilon''}{\epsilon'} \right)^2} + 1 \right) \right]^{\frac{1}{2}} \quad (9)$$

which defines the movement of the phase front, and it is related to the propagation wavelength it accounts for the distance between wave fronts which are separated in time by one period of oscillation. Thus, for moderate lossy materials, the detectable size of targets will be limited by the spatial resolution  $\lambda = \frac{2\pi}{\beta}$ .

For lossy materials, the wavelength and the phase constant are frequency-dependent, and that will be a distortion of the pulse as it propagates through the host media. To operate in these dispersive media, GPR systems use different kinds of modulations which create propagating waves in a range of frequencies. The velocity of this packet of waves is called group velocity. In order to detect a target in a lossy medium, several different waves are transmitted, and they arrive at the target location at different times and with different phase angles. Even though this solution increases the depth range and the resolution, it also influences the waveform of the detected signal, making it more difficult to interpret the results.

The polarization of plane waves is another relevant parameter for the performance of GPR. Considering that the field intensity of a uniform plane wave has a direction in space which may either be constant or time-varying, the polarization corresponds to the curve traced by the field vector in space as it propagates. The trace can depict a straight line, an ellipse, or a circle, corresponding to linear, elliptical, and circular polarization, respectively. Key systems of the GPR equipment are the transmitting/receiving antennas. In general, each antenna can transmit/receive electromagnetic waves with a certain polarization, and both should be oriented properly to maximize the detected signal. However, in practice, the material property can change the polarization of a plane waves, and this will result in some losses due to polarization mismatches.

The reflection and refraction of the EM wave are the main phenomena involved in the detection of any target. In a GPR survey (Fig. 1), an incident wave propagating from one medium (typically air) reaches a different medium (structure under test or host). A fraction of the wave is then transmitted (or refracted) to the host, while another fraction is reflected. The proportion of transmitted energy can be calculated by accounting for the intrinsic impedance of the air and the host medium, which can be defined as:

$$\eta = \left[ \frac{-j\omega\mu}{\sigma - j\omega\varepsilon} \right]^{\frac{1}{2}} \quad (10)$$

and the angle of incidence of the EM wave to the surface of the host medium. Similar reflection/refraction phenomena happen in the host medium when the incoming wave impinges on a flaw/target of different constitutive parameters to those of the host, as well as a certain attenuation due to the lossy properties of the host. As a result, one should assess the material electrical properties of the surveyed media to predict the amount of energy required to detect any target of a minimum size at a maximum depth.

## Material Properties

As it can be inferred from the previous section, a major step in any successful detection depends on the right estimation of the electrical properties of the structure under test. Interactions between electromagnetic fields and matter in GPR have to be

considered from a macroscopic point of view. Nevertheless, the constitutive parameters of the media are derived from a microscopic point of view. Matter consists mainly of atoms or molecules which in turn contains positive and negative charges. Charges that can travel through the media are called free charges, and those which cannot move freely (mainly because of a strong interaction with other of opposite sign) are called bound charges. In general, materials can be characterized by the predominant response of the charges when an electromagnetic field is applied to the structure. As the GPR technique is generally applied to lossy dielectric structures, the relevant physical responses are the following:

- Polarization. In this response, the bound charges respond to the external electric field by creating small dipoles, which are polarized in the electric field direction. The permittivity is mostly a measure of the material polarization.
- Conduction. Free charges move through the matter as a response to the external EM field, a process in which the material absorbs energy as a result of microscopic interactions. The result is dissipation of power of the electromagnetic wave due to conductivity.

It should be emphasized that electrical properties are related to (a) frequency, (b) position, (c) intensity, and (d) orientation of the driving electromagnetic field. Considering (a), some dielectrics exhibit dipolar polarization or have molecules which are permanent dipoles. When an external low-frequency driving electric field is applied to this material, the charges in these molecules have sufficient time to change their direction. The dielectric permittivity is then frequency-dependent, and the material is called dispersive. The time needed for dipole orientation is called the Debye relaxation time, and the wave propagation in this medium is nonlinear in frequency. Examples of these materials are those with water content, as, for instance, concrete.

Regarding the position, when the medium's behavior varies when a uniform driving field is applied at different locations, this medium is called inhomogeneous. In this case constitutive parameters vary with position in space. In open-field GPR surveys, this inhomogeneous host property could lead to change of the maximum depth of analysis at different locations. In the case of the intensity, for most applications, the polarization of the medium responds linearly in relation to the amplitude of the incident field. Therefore, in this chapter, we consider linear materials. In the same line, there are materials that are not equally polarized for different orientations of the driving field (d), but they rarely have to be considered in GPR applications. Table 1 illustrates the electrical properties of typical materials for GPR at a frequency of 100 MHz.

Following the guidelines of the previous section, one can clearly infer that the EM waves suffer less attenuation in some materials such as air or clay than in others like concrete or asphalt, due to their lower conductivity. In general, the performance of a GPR assessment will be better in a host composed of simple material (i.e., lossless, linear, isotropic, and homogeneous) with targets of high electrical contrast (i.e.,  $\epsilon$  and  $\mu$  much different than those of the host medium), where high reflections occur (Daniels 2004; Jol 2008).

**Table 1** Electrical parameters of materials

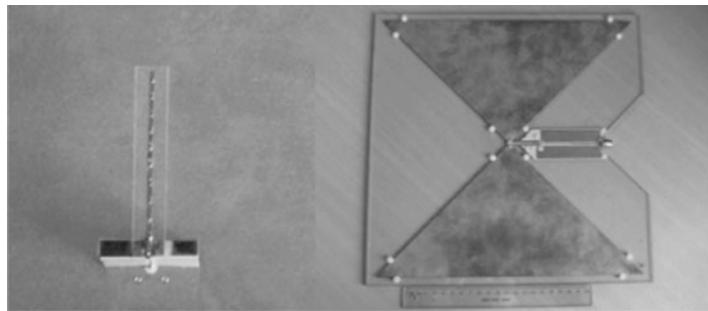
Material	Permittivity range	Conductivity
Air	1	0
Asphalt	2–4	$10^{-2};10^{-1}$
Clay	2–6	$10^{-1};10^0$
Concrete dry	4–10	$10^{-3};10^{-2}$
Epoxy	3–4,5	$2.10^{-3}$
Glass	4–7	$10^{-10}$
Granite	5	$10^{-8};10^{-6}$
Limestone	7	$10^{-8};10^{-6}$
Mica	6	$10^{-15}$
Polypropylene	2,5–2,6	$10^{13}$
Polystyrene	2,56	$10^{-14}$
Polyethylene	2,26	$10^{-13};10^{-17}$
Porcelain	6–8	$10^{-10};10^{-12}$
PVC	3,39–4,52	$10^{13}$

## Antennas

Antennas are responsible for the generation of the waves at the appropriate power levels and range of frequencies. Their description is particularly useful because it provides a review of the possible configurations of the GPR system, according to the specific intended application. The role of antennas is to transduce energy, generated in a radio-frequency circuitry, from a guided medium to an unbounded media. In recent years, the demand for the innovation in GPR antennas has been rapidly increasing due to the wider scope of applications covered by GPR. In the sense of the performance parameters, a high-quality antenna or array of antennas is essential for the GPR resolution, penetration depth, and target detection. But additional requirements are for the practical application of GPR to ease the survey costs: low-profile, pure linear-polarization, and compact size, among others.

For instance, hydrological and geological applications often demand a high penetration depth for a low resolution, while other civil applications such as concrete inspection require opposite characteristics. Hydrological applications can afford larger size antennas which achieve deeper penetration by operating at a lower range of microwave frequencies. On the other hand, concrete structures require compact size antennas which have higher resolutions but lower penetration depth. In this way, the choice of antenna establishes a tradeoff between resolution and penetration depth according to the central frequency in which it operates.

Basically, one can consider an antenna as a two-port transducer system characterized by a transfer function that can be frequency-dependent and has a nonlinear phase response. In this case, the waveform that originates in the electronic circuitry will be significantly distorted, and the radiated pulse can be inappropriate for the postprocessing stage. This fact can be critical for an impulse wave radar, whereas in frequency-modulated or synthesized GPR, a simple calibration will fix the problem.



**Fig. 3** GPR antennas (loaded dipole and bowtie)

In this sense, the choice of the system domain will be again a trade-off. Time-domain radar antenna may produce ringing, signified by means of oscillations which could mask signals reflected from the targets. Frequency-domain radars will require longer survey times, and it will be unappropriate for real-time applications.

In general, the antenna characteristics are influenced by the surrounding or nearby media. The efficiency of a GPR antenna can be substantially reduced when it operates in close proximity to a dispersive medium, not only because of the impedance mismatching but also for more demanding operating power which, in applications at isolated locations, may require larger batteries. Also, media in close proximity can affect the operational bandwidth of antennas. In general, most GPR antennas are wide or ultra-wideband antennas in terms of the operational frequency range. In this sense, the higher the bandwidth the lower the efficiency, in terms of radiated power of the antenna. For this reason, the choice of antennas depends on the application. Borehole radars use narrower bandwidth antennas (such as dipole or loop antennas), while impulse radars are made of wideband or ultra-wideband antennas (bowtie, loaded bowtie, loaded dipole, spiral, horn, and Vivaldi types). Figure 3 illustrates some GPR antennas.

Next we summarize the key antenna parameters in conjunction with GPR equipment (Balanis 2005). First, the bandwidth of the antenna is described as the range of frequencies over which the antenna can maintain appropriate radiation characteristics. Considering GPR systems, the bandwidth is one of the determining parameters used to decide upon an antenna type. For instance, many antenna types have very narrow bandwidths and cannot be used for transmitting a pulse with wide frequency content. Bandwidth is typically quoted in terms of the input impedance, although that is not the only criterion to be taken into account. A mismatch of the input impedance is due to the reflected waves within the antenna, which increase the voltage standing wave ratio (VSWR).

Second, the polarization of the antenna plays a major role in the survey. Commercial GPR antennas can be found with linear and circular/elliptical polarization according to the application to avoid polarization mismatch. As long as the polarization of the radiated wave can be affected by reflection/scattering phenomena or even by propagation in heterogeneous soils, the use of elliptical polarized antenna seems advisable because they can receive both linear and elliptical polarized waves.

However, the radiated energy at a particular polarization is higher in a linear polarized antenna and may not be able to receive so efficiently as elliptically or circularly polarized antennas. Therefore, for targets oriented in a particular way (e.g., steel bars in concrete), the use of linear antennas could occasionally make sense. If a priori information on how the target is oriented is not available, elliptically polarized antennas are preferred. A good example of practical antennas with elliptical polarization is the log-spiral (Fig. 4). In this configuration the transmitting antenna radiates two patterns that are orthogonal at any location. Using cross-polarized antennas, it is possible to detect simultaneously the target reflection in two orthogonal polarizations (Nicolaescu 2013), and some detected signal will be received no matter what the orientation of the target.

Additional attention should be paid to the directivity and gain of the antenna. The radiated power of any GPR antenna varies for different directions in space, giving directive properties to the antenna. The directivity is mathematically defined as the ratio between the radiation intensity in a particular direction and averaged radiation intensity, disregarding any losses that may have occurred in the antenna itself. If losses in the antenna are considered, by accounting for the input power available to the antenna, one defines the antenna gain or power gain. If no mention is made of the direction associated with directivity, the maximum directivity or maximum gain should be understood, and manufacturers should clarify in the characteristic sheet what is the direction of maximum radiation.

Finally, as a parameter which is particularly interesting for GPR purposes, we mention the antenna footprint. It is defined by the illuminated area in the horizontal plane at a given distance from the antenna in the structure under test. This energy projection is substantially influenced the frequency, material properties, and antenna elevation. In general, when the material permittivity is higher, the footprint is more

**Fig. 4** Log-spiral antenna



compact, while for lower permittivities, the footprints become larger. For closely spaced targets, a small footprint is desired since it is important to distinguish between buried targets. In addition, a small footprint is useful to avoid surface reflections and clutter. For optimal detection the footprint should comprise the size of the target, or detection will be compromised. If the transmitted power of the antenna is spread across an area much larger than the target, then the signal backscattered will be much weaker than the signal reflected by the surface. The footprint has an elliptical shape as illustrated in Fig. 5, with an approximate size (Neal 2004):

$$A = \frac{\lambda}{4} + \frac{D}{\sqrt{\epsilon_r - 1}} \quad (11)$$

where  $A$  is the approximate long dimension radius of the footprint,  $\lambda$  is the wavelength of the EM wave, and  $D$  is the depth. The optimal size and shape of the footprint can vary depending on the application. For instance, if the GPR is intended to detect long rebars in concrete structures, the footprint should have a longer ellipse shape, while small targets such as cracks require circular footprints.

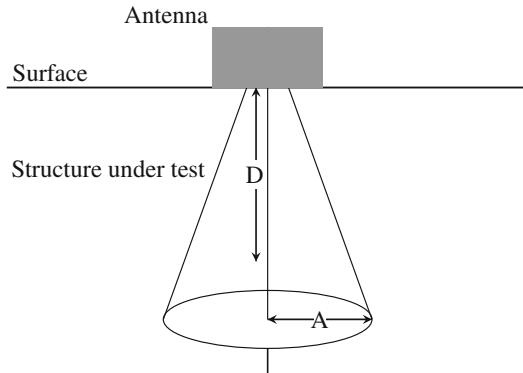
Table 2 shows recent works on GPR considering different antenna types and applications.

## System Specification

GPR systems can offer different configurations according to their system components. The system configuration directly affects the performance of the NDT assessment. Since often some prior information about the structure under test and the target characteristics is known, the system specification plays a major role in the success of the assessment.

Considering the particular application, GPR can perform one or more of the following tasks: (a) the examination of a given volume in order to detect potential targets, (b) imaging of the target, and (c) classification of the target characteristics. The criteria that should be evaluated for those purposes are the reliability of

**Fig. 5** Antenna footprint



**Table 2** GPR applications

Reference	Application	Antenna type
Lestari et al. (2010), Sagnard and Rejiba (2011), Giannakis et al. (2016)	Lossy dielectrics	Bowtie
Ebihara et al. (2012), Qin and Xie (2016), Hfinghoff and Overmeyer (2013)	Water leakage, metal ore detection	Loaded dipole
De Coster et al. (2016), Almeida et al. (2016), Shao et al. (2014)	Forensic survey	Horn
Nicolaescu (2013), Liu et al. (2015)	Landmine, rebar detection	Spiral
Alkhalifeh et al. (2016), Guan et al. (2017)	Lossy, dispersive media	Vivaldi

detection, the accuracy of the predicted position/features of the target, the resolution of the survey, and the immunity to environmental noise.

The reliability of detection should account for the detection range considering the expected features of the target and the host medium's electrical parameters, including possible nonhomogeneities. To this end, the system must be able to distinguish clutter and noise signals, reflected signals from the surface, and real target signals to avoid false positives.

Target detection and classification should be done as accurately as possible and should include information about its position, size, shape, orientation, and electrical properties. As stated earlier, the host medium and target should have different properties, so the incident wave can generate a strong reflected wave to be detected by the receiving antenna. In a GPR survey, targets must be resolved from other targets by one or more spatial coordinates. This is called spatial resolution and measures the ability to distinguish (or discriminate) reflections from closely located targets.

Finally, unintentional electronic interferences can preclude or severely limit the use of GPR. For this reason, GPR systems should be less susceptible (or have immunity) to environmental radio frequency interferences. The RF immunity measures the ability of a GPR to perform its functions in the presence of undesired signals (noise). The GPR can be classified by its frequency of operation, and to minimize interference from external EM sources, GPR manufacturers are authorized to use specific parts of the EM spectrum by national government agencies.

Therefore, the first choice to be made should be the frequency of the survey, which drives relevant features such as penetration depth and resolution. In fact, there is a trade-off in defining those two considerations. Since the penetration depth is directly related to the attenuation, one can infer from Eq. 8 that higher frequencies will survey shallower depths than low frequencies. On the other hand, from Eq. 9, a higher frequency  $f$  means a lower wavelength  $\lambda$ , which also implies a lower horizontal resolution (footprint) of the GPR system as predicted in Eq. 11 (Jol 2008). In summary, the frequency selection drives important parameters in the GPR survey including the sampling criteria in time and space. This is controlled by the

acquisition parameters of the equipment, such as the time sampling and the space between readings. To avoid aliasing, the Nyquist theorem provides maximum time and spatial intervals for a frequency  $f$  as:

$$\Delta t \leq \frac{1}{2f} \quad (12)$$

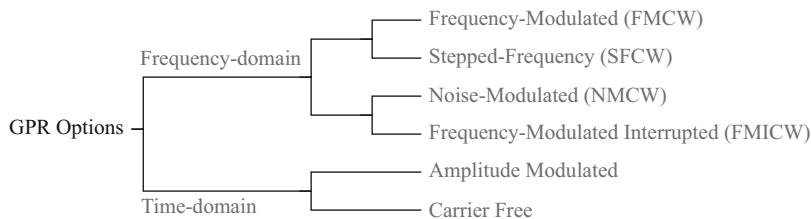
$$\Delta x \leq \frac{\lambda}{2} \quad (13)$$

In terms of vertical resolution (that is, the ability to discriminate stacked targets), the bandwidth of the system is relevant, and the radiated waveform is another choice to be made. As it was mentioned, any EM pulse following a path through different media undergoes changes in its original waveform. These modifications will impact the GPR performance, and the waveform definition is a crucial factor. Some aspects of the waveform should be considered including (a) signal to noise ratio, (b) avoidance of false-positive detection, (c) resolution required, and (d) electromagnetic compatibility (EMC) issues. It is important to note that those considerations are often in contradiction.

GPR can operate with several modulation techniques. The most common are as follows:

- Continuous-wave (CW) radar utilizes stationary waveforms, usually sinusoids, instead of pulsed waveforms. This is an important issue for measuring the range since there is no reference time between transmission and reception. In order to solve this problem, some frequency modulation (FMCW) of the waveform can be used. To avoid interruptions of the CW, two antennas are used in a transmitter/receiver configuration.
- Pulse-compression radar uses pulse waveforms coded with some modulation aimed to improve the signal bandwidth and provide similar vertical resolution to that of short (in terms of temporal duration) pulses. In general, the modulation is frequency or phase-based. There are several types of pulse-compression techniques for radar, with the chirp and stepped-frequency methods being the most common.
- Synthetic-aperture radar (SAR) maintains phase coherence for a transmitted pulse at a certain range of directions called antenna aperture. By depicting the phase information of the received signal, a high-resolution image provides the target features. Large antenna apertures can be synthesized as a superposition of measurements made by small aperture antennas.
- Impulse radars were developed to emit very short pulses of electromagnetic waves with high peak power. For this reason the technique is often called short-pulse radar, impulse radar, or ultra-wideband (UWB) radar.

Figure 6 shows the most common options for GPR systems either in time or frequency domain. Presently, most of the commercial systems use time-domain modulation, but the employment of frequency-domain systems is growing fast due to the fact that they require less power and are more robust in the presence of noise (Pieraccini 2018; Leckebusch 2011).



**Fig. 6** GPR systems according to the waveform modulation

Frequency-domain GPRs are based on band-limited signals. The most basic one is the continuous wave (CW) GPR, which transmits a signal of infinite duration in time and consequently narrowband in frequency, as a continuous sine wave, and receives it simultaneously. In this configuration, it is possible to detect buried targets, but it is not possible to resolve range since the signal do not change with range. For this reason CW GPRs are not practical, and the signal bandwidth is widened by modulating it in several ways. A first choice is to employ amplitude modulation, in which the continuous wave is modified by considering amplitudes 1 and 0 (on and off states) at different times during the inspection. This option is commonly named pulsed GPR. Since the transmission and reception times are well defined in the equipment, it is possible to associate then with the target range. However, for a greater resolution, it is desirable to control the power spectral density of the pulse.

Another approach for the modulation is to add more frequencies by increasing or decreasing the oscillation in the waveform (FMCW). Considering that response from the target is frequency-dependent, the survey of the host is improved. However, this type of GPR systems suffers from interference issues, and a different method can be employed in the transmitted frequency scans a certain interval of frequencies at different frequency steps, covering thus the frequency range in a discrete way. This technique is called stepped frequency continuous wave (SFCW). Also, the frequency step avoids phase ambiguity by measuring the phase difference between returned signals at each frequency. In addition, the hardware associated with SFCW is simpler than that employed in FMCW. A variation of the SFCW technique involves gating the timing of the transmitter and receiver circuits. It is possible to use other options including noise-modulated continuous waveform (NMCW) (Reeves 2014) and frequency-modulated interrupted continuous wave (FMICW) (Raimundo et al. 2014).

On the other hand, GPR systems that radiate a time-limited waveform and are aimed at detecting changes in its amplitude are known as impulse GPRs. In these systems, the bandwidth is directly related to the pulse width. Shorter-time pulses require larger bandwidths and consequently wideband antennas. The impulse GPRs is still prevalent over the FMCW and SFCW options due to the cost of RF components and the hardware simplicity. However, they present some drawbacks including undesirable ringing, limited resolution, and phase group distortion.

Considering the SNR, both FMCW and SFCW have lower noise figures and require less power to detect targets than impulse GPRs (Nicolaescu 2013). In terms

of EMC issues, GPRs should avoid to send any spurious EM signals into the air which could interfere with other electronic systems. To this end, it is mandatory to comply with EMC standards ruled by national agencies such as the Federal Communications Commission (FCC) in the United States and the European Telecommunications Standards Institute (ETSI) in Europe.

Another classification of GPR systems can be made in terms of the coupling to the structure under test as air-launched or ground-coupled. Antennas can be touching or even inside the host medium in ground-coupled GPRs, or they can be located above it in the air-coupled GPRs. This choice will affect important features in the GPR assessment such as the footprint and the coupling signal, in amplitude and waveform, between the radiated pulse and that propagating in the host medium (Nicolaescu 2013).

Table 3 shows some GPR commercial systems according to their coupling to the structure under test. The efficiency and input impedance of a GPR antenna can be substantially reduced when it operates close to a dispersive medium, requiring more power to operate. The specification of how close the antenna can operate from the structure under test has to be considered in this case. In general, ground-coupled GPR can transmit/receive more power to the host medium, which results in clearer data and greater depth. But the eventual surface coupling and antenna ringing can result in difficult in obtaining accurate predictions without appropriate signal processing. Air-coupled GPR measurements, on the other hand, must contend with the initial reflection from the air-ground surface, and less energy is thus used in the survey, but since the antennas are not in contact with the ground, data can be collected at a much higher speed without damaging the antennas.

Once the desired waveform is generated, a transmitting antenna should radiate it adequately. To this end, the antenna should match the spectrum of the selected

**Table 3** Commercial GPR systems

Manufacturer	Frequency (MHz)	Coupling
CyTerra	Not specified	Air
Ditch Witch	250, 700	Ground
Easyrad	20–400	Air
Easyrad	500–2000	Ground
Geoscanners	210,307,380, 480, 675,1000, 1500, 2000	Ground
Logis-Geotech	250–700, 90–2500	Ground
Logis-Geotech	400–2500	Air
GSSI	270, 400, 900, 2000	Ground
GSSI	1600, 2600	Air
IDS GeoRadar	400–900, 200–400, 600–900	Air
IDS GeoRadar	2000	Ground
Mala Geoscience	200, 250, 1200, 1600, 2300	Ground
Terraplus	200, 450, 500, 750, 1000, 1600, 2300, 200	Ground
US radar	250, 500, 1000, 1500, 2000, 2500	Ground
3D-Radar	200–3000	Ground

waveform. In this sense, a priori information about the test scenario is relevant to select an antenna structure less prone to undesired reflections, noise, and interference signals. Planar PCB antennas are mainly used in ground-coupled GPRs for their non-dispersive characteristics. In terms of laboratories and controlled surveys, air-coupled GPRs using horn antennas present remarkable advantages because of their narrow beamwidths and high gains over a wider frequency range.

The antenna structure can impact the GPR readings imposing practical trade-offs. Available GPR equipment shows a wide range of antennas with different radiation patterns and sizes. Special attention has to be paid to the choice of the spatial interval of measurement  $\Delta x$ , which could shorten the time of the survey at the cost of spatial resolution. A careful choice can avoid data volume exaggeration or a spatially aliased response, but, in case of unfortunate decisions, it will jeopardize the predictions achieved at the postprocessing stage.

Recent years have brought new designs involving arrays of antennas, rather than the classical transmitter/receiver configuration. For bistatic (two antennas) or multistatic (array of antennas) cases, the distance between antennas should also be considered in the system specification. This feature will change the radiation pattern of the antenna and its coupling with the host medium, and it has to be made in accordance with the size of the target. The use of multiple-input multiple-output (MIMO) array antenna technology has been proven to substantially improve the target detection resolution (Zeng et al. 2015), and it is expected to grow in the near future.

In summary, the system specification defines how the GPR operates based on the target characteristics and the structure under test. Important choices must be made including the system bandwidth and the power employed in the assessment. The definitions can improve the likelihood of detection producing better underground images. However, inaccurate electromagnetic modeling of the scenario can cause misinterpretation in direct readings, leading to wrong predictions in the actual location. In order to improve the data interpretation for GPR user, postprocessing support tools (PST) are necessary. They substantially improve the assessment by eliminating the noise and other potential sources of errors in the measurements.

## A practical application: GPR detection in concrete slabs

This section discusses the application of feature selection based algorithms for detection of various targets in concrete slabs. In this case the feature selection is a type of pre-processing, using some pre-defined cost function, as the classes separation margin. The algorithms considered here are Relief and Simba as in (Gilad-Bachrach et al. 2004). The Relief algorithm is based on the statistical relevance of the features, while Simba is based on the concept of the separation hypothesis-margin between two classes.

From the reflected wave, 15 features were computed to be selected.

1. Delay of the first reflection (delay) (Caorsi and Cevini 2005).
2. Maximum amplitude ( $\max(G_t)$ ) (Caorsi and Cevini 2005).

3. Reflected wave mean (mean( $G_t$ )).
4. Reflected wave standard deviation (std( $G_t$ )).
5. Mean of the wave derivative (mean( $dG_t$ )).
6. Standard deviation of the wave derivative (std( $dG_t$ )).
7. Maximum amplitude of the Fourier transform (mFFT).
8. Energy in six different bands ( $B_1, B_2, \dots, B_6$ ) of the Fourier transform.
9. Frequency of maximum amplitude of the Fourier transform (fmFFT).
10. Signal energy (sen( $G_t$ )).

The six most commonly selected features (Queiroz et al. 2013) were *mean( $G_t$ )*, *mFFT*, *fmFFT*, *sen( $G_t$ )*, *std( $dG_t$ )*, and *Delay*.

The feature reduction avoids large structures to data set modeling in both regression and classification problems. In addition, a smaller set of features may expose models to less noisy data, and hence, it may improve classification accuracy. Other advantages are a reduction of computational resources needed to store data and calculated metrics from samples. And models with fewer features tend to facilitate the development of more explainable models because the resultant features can be justified by known information which relates samples and their labels.

To do so, the feature selection methods must have the ability to deal with large data sets with many samples and features. Depending on the selected strategy to feature extraction, which replaces the original data by some of their metrics for evaluation, the new representation of each sample should not vary too much due to noise. If this is not the case, the feature selection process may be affected by metrics intolerant to noise. The final subset of features may greatly vary besides requiring a consensus response and a more complex data pre-processing step.

Therefore, feature selection methods must be applied more than once. Another problem related to filtering methods for feature selection is the requirement of having all classes well represented in the training set. As the methods addressed in this work are based on hypothesis-margin, the samples near the margin should be present in the set considered in the ranking of features. The metric and strategy used to compare the combinations of features have considerable impact on the quality of results and the number of evaluations. As a consequence, different filter methods can reach distinct results requiring similar time for evaluation. Therefore, in light of the problems discussed above, the final selected features are not guaranteed to be the best subset.

We nextly present the data processing for this application. Concrete has been the most commonly used building material because it is durable, cheap, and readily available and can be cast into almost any shape. The main components of concrete are the hydrated cement paste, aggregates, water, and the transition zone.

The main constituent of concrete is cement, which is made up of calcium, silica, alumina, and iron oxide. The aggregates occupy 60–80% of the concrete's volume. They are usually stronger than cement and therefore play a role as a filler (Mehta 1986).

Concrete has many voids that are usually filled with water. The water can be classified depending on the type of void and degree of firmness with which the water

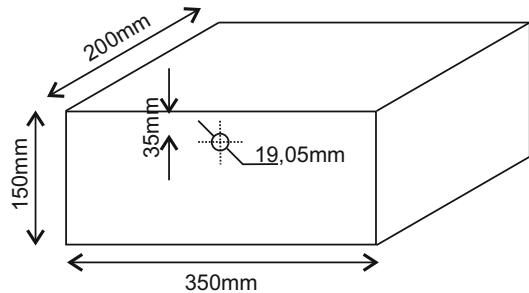
is held in these voids. The water, which is responsible for the electrical conductivity of concrete, is called free water and is present in voids greater than 50 nm in size.

The transition zone, the region that exists between the hydrated cement paste and the aggregate, is very thin, with a thickness on the order of 10–50  $\mu\text{m}$ . The transition zone is important in that it is the weakest zone in the concrete and thus influences its stiffness and durability.

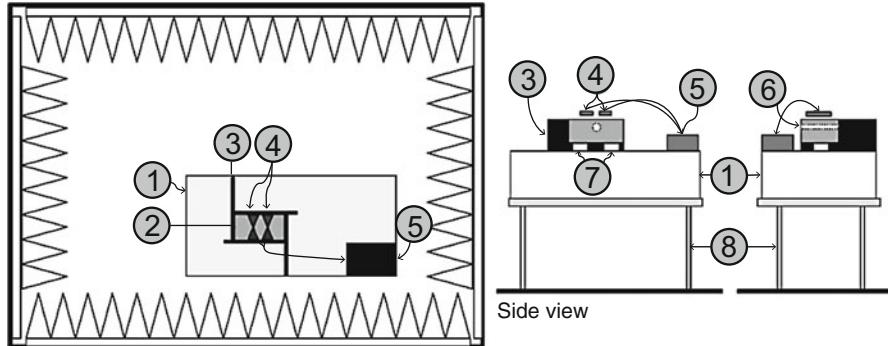
The concrete used in the experiments described here has dimensions illustrated in Fig. 7. Four slabs were fabricated to meet the characteristics found in Bykztrk et al. (2006) and Al-Qadi et al. (1995). The mixture has the following characteristics: (I) cement Portland type I, (II) water/cement ratio 0.60, and (III) cement/sand ratio 1:2.25. The inclusions are at a depth of 35 mm from the surface of the concrete slab and have diameters of 19.05 mm as illustrated in Fig. 7.

In order to obtain experimental results, a GPR survey was conducted on the four concrete slabs with the following characteristics: (A) without inclusion, (B) with metal inclusion, (C) with PVC inclusion, and (D) water-filled inclusion. The survey was performed in a semi-anechoic chamber to avoid noise in the experiment according to the test setup illustrated in Fig. 8.

**Fig. 7** Concrete slab dimensions



Upper view



**Fig. 8** Schematic and details of the assembly placed in a semi-anechoic chamber. (1) Polystyrene layer, (2) concrete block, (3) absorber plate, (4) bowtie antennas, (5) GPR control unit, (6) inclusion, (7) Polystyrene blocks, and (8) wood table

The GPR equipment used in this work has minimum configuration consisting of the following devices grouped and used together:

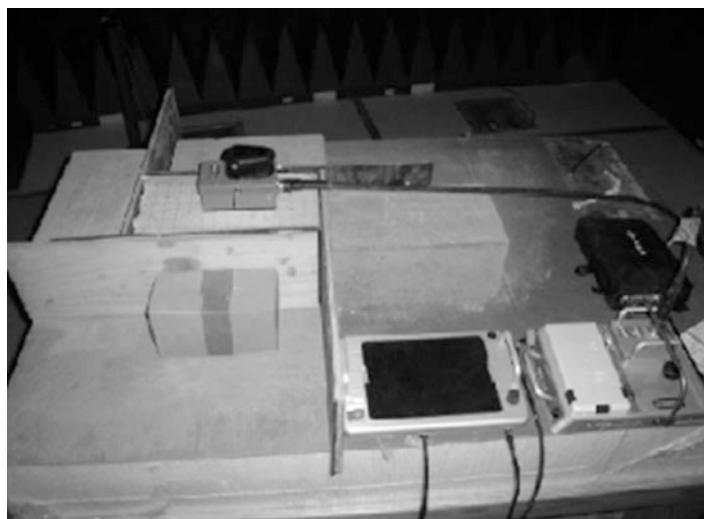
- (a) 1.6 GHz shielded antennas or 2.3 GHz shielded antennas
- (b) Control unit.
- (c) Monitor XV-11 with 1.2 m cable.
- (d) Cable X3M 4 m.
- (e) Battery li-ion 11.1 V/Ah.

Prior to the GPR assessment of the concrete in the chamber, test setup ambient levels (i.e., all equipment energized except the GPR equipment) were performed to verify a 6 dB or more below the level required for testing. The ambient measurements were performed using vertical polarization of a log-periodic antenna.

The GPR equipment and its cable harness, along with the concrete slabs, lie on an insulated support 1000 mm (+/-50 mm) above the floor of the test chamber. The dielectric relative constant of the insulated support were less than 1.4. The groundplane was not used for this experiment. The battery was located under the test bench. The equipment was at least 1000 mm from the chamber walls. No part of the GPR equipment was closer than 250 mm to the floor.

The GPR operated under typical loading and other conditions as in a regular survey to detect cracks and inclusions. The cables provided by the GPR equipment to perform the tests were shielded. Some cables were excessively long and were not bulky and were bundled at the approximate center of the cable forming bundles 300–400 mm in length.

Figure 9 illustrates the experiment in the semi-anechoic chamber. The climatic test conditions are temperature  $23 \pm 5^\circ\text{C}$  and relative humidity between 20% and 80%.



**Fig. 9** Experiment performed in a semi-anechoic chamber

Each concrete block was probed on its top and bottom sides. Therefore, 324 samples were collected with each antenna and 108 samples for each block inclusion (air, conductor, or PVC + water). Since 12 samples of each block were taken over the inclusion, only these samples were considered in the depth regression algorithm with half of them from top and the other half from the bottom of the blocks. The radius estimation was not experimented with the real data set since all inclusions in the concrete blocks had the same diameter.

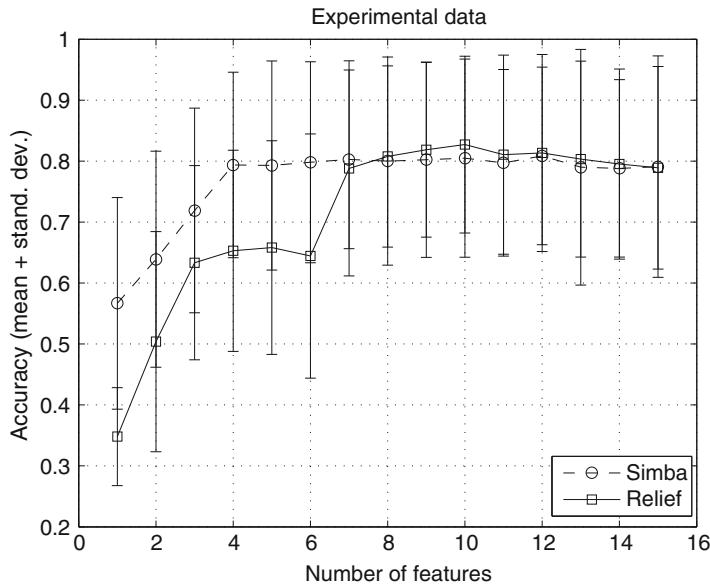
Each A-scan from both antennas had the first echo excluded from data for models experiments. For the 1.6 GHz antenna, only samples from the 80th to the 240th position were considered. For the 2.3 GHz antenna, samples were selected from the 54th to the 200th position. Before submitting these samples to classification and regression models, each data set was subject to feature extraction as in Queiroz et al. (2013) and after that to Simba and Relief models to rank features by their capacity of class discrimination. The ranking results (from most to least important) for each antenna are as follows:

- 1.6 GHz antenna
  - Simba:  $f$  mFFT,  $\max(W)$ ,  $\text{mean}(dW)$ ,  $\text{std}(dW)$ , mFFT, B6, SignalEnergy, B2,  $\text{std}(W)$ , B3, B5, B1, B4,  $\text{mean}(W)$ , Delay.
  - Relief:  $f$  mFFT,  $\max(W)$ ,  $\text{std}(dW)$ , B3, B5, B2, mFFT, SignalEnergy,  $\text{mean}(dW)$ ,  $\text{std}(W)$ , Delay, B6, B1,  $\text{mean}(W)$ , B4.
- 2.3 GHz antenna
  - Simba:  $f$  mFFT,  $\max(W)$ ,  $\text{mean}(dW)$ ,  $\text{std}(dW)$ , mFFT, B6, SignalEnergy, B2,  $\text{std}(W)$ , B3, B5, B1, B4,  $\text{mean}(W)$ , Delay.
  - Relief:  $f$  mFFT,  $\max(W)$ ,  $\text{std}(dW)$ , B3, B5, B2, mFFT, SignalEnergy,  $\text{mean}(dW)$ ,  $\text{std}(W)$ , Delay, B6, B1,  $\text{mean}(W)$ , B4.

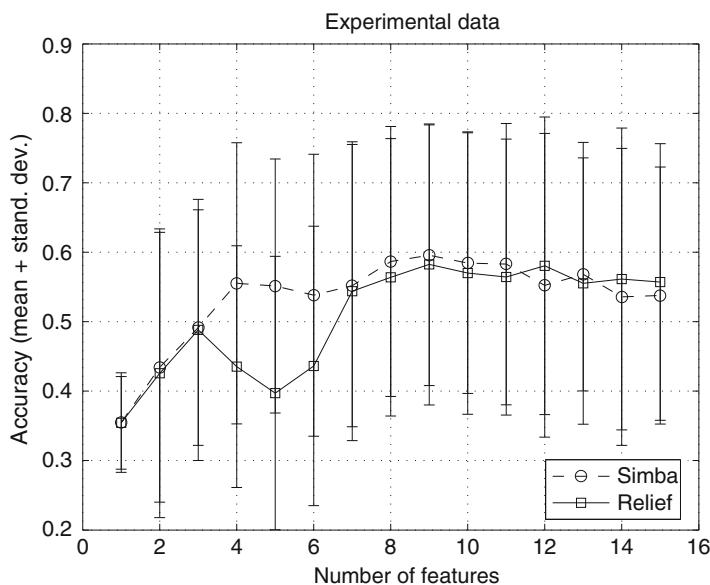
Each of these features ranking was important to generate results to be compared to simulated data sets. As in the latter, no pre-processing of data was done before sample submission to models.

The  $k = 3$ -NN for inclusion classification model used Squared Euclidean distance. Its accuracy was computed adopting  $n = 50$ -fold cross-validation method adopting *cvppartition* in MATLAB with its default values. The corresponding results are registered in Fig. 10 for 1.6 GHz antenna and in Fig. 11 for 2.3 GHz antenna. Higher accuracies with fewer features were attained using the sequences obtained by Simba. But in both data sets, the accuracies were smaller than expected. The results are seen to be affected by real data problems like noise, clutter, and equipment instabilities such as signal shifting. Clearly, more care should be taken in pre-processing steps to improve the quality of signals submitted to modeling steps.

The feature values extracted from both data sets permitted stable results despite the apparent higher standard deviation in Fig. 11 because of the magnification done to improve visualization. The features added to this model indicated the existence of a small group of six features which should be considered to achieve smaller percent errors. On the other hand, in both cases, the Delay feature, which showed constant values for all samples, had a significant impact on the chosen model, causing bad



**Fig. 10** Accuracy results for  $k$ -NN with experimental results for the 1.6 GHz antenna used to classify types of inclusions



**Fig. 11** Accuracy results for  $k$ -NN with experimental results for the 2.3 GHz antenna used to classify types of inclusions

conditioned matrix weights. The first models without considering the Delay feature had significant percent error for depth estimation. On the other hand, the subsequent models with Delay and other new features could not be trained satisfactory. As a consequence, these features should be disregarded until a new function for first delay arrival can effectively extract the corresponding information for samples discrimination.

As could be seen, a few extracted features are not robust to noise contamination and need some pre-processing to be reasonably evaluated. Some models were accordingly developed to reduce signal contamination, reduce the data formatting steps, or in some way overcome the large number of features and samples from A-scans. But the feature extraction process still has to be improved to avoid constant values, to gain a small general set of characteristics, and to reduce the pre-processing steps needed. In other words, a set of features must be developed to efficiently capture the information in which it is expected to find the elements capable of discriminating between materials and explaining the phenomena involved in this inverse problem. More informative and less complex models were obtained by the use of specific models to each inverse problem and trying to include some A-scans features that visually distinguish the materials and indicate inclusions.

---

## Postprocessing Support Tools (PSTs)

In addition to the concepts presented in the previous section, the on-site GPR measurements require a postprocessing stage to achieve the major goal of any prospection: a prediction of geometrical features and physical properties of any embedded object in the host. In this sense, it is key to realizing the many factors not presented here and which affect the acquisition of real data: the non-homogeneities of the host medium, the lack of accurate material descriptions in actual scenarios including varying environmental conditions, the superposition of different scattering waves coming from multiple flaws, the roughness of the interface of the host medium, the eventual electromagnetic compatibility issues coming from external sources, the changes in the footprint of the antenna due to the presence of near objects, and the electrical noise of the equipment. These phenomena are a short list of major issues obviously neglected in any basic introduction. As a result, electromagnetic wave propagation and scattering effects are so complex that any signal or image depicted from raw data provides only limited information, and PSTs constitute the only way to enhance the system performance providing concise predictions in real-world applications.

In most of the cases, GPR manufacturers provide proprietary software which consider the particular system characteristics of the model and aimed not only to provide survey maps of the site (generically named datagrams) but also to get a real-time check of the quality of measurements. In this sense, any malfunction of the equipment (e.g., defective triggering pulse, inappropriate connections of cables) is detected by this software, saving thus efforts and time in prospection.

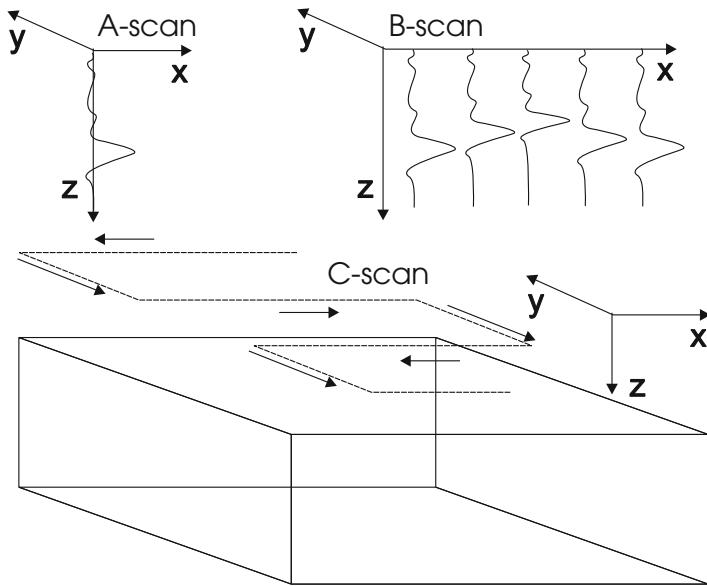
Additionally, a subsequent secondary set of PSTs sold separately are usually applied for an accurate reporting of the GPR survey. In this sense, it is useful to emphasize that generic codes which neglect the basic electromagnetic phenomena of section “[Fundamental Theory of GPR](#)” will be of limited utility in most cases, and it is only by taking into account these phenomena that the predictions start to be effective. In any case, there is still on-going research aimed to combine the latest techniques coming from signal and image processing theory with the specific electromagnetic properties of a GPR detection ([Rodriguez et al. 2015](#)). Inclusion of advanced modeling features into the PSTs increases successful rates, contributing thus to closing the so-called inverse problem gap which presently is the bottleneck in a wider use of GPRs in generic applications. Table 4 shows a short list of GPR software.

In short, the available information from GPR equipment after on-site measurements is far from being ready to provide the location and characteristics of the target. Auxiliary software tools may improve the interpretation of the collected data, thus reducing substantially the time and cost of the GPR detection. The objectives of those tools are to detect, locate, and define target’s geometrical and physical properties. While basic techniques are focused on the first two goals, more advanced techniques deal with all objectives at once. In this sense, the starting point to classify PSTs is based on how measurements are gathered. According to the spatial acquisition of the survey, and recalling that a time-signal will be measured for any on-site location, the available information can be processed separately at each location (called A-scan), or it may grouped along lines (B-scans) and surfaces (C-scans) (Fig. 12).

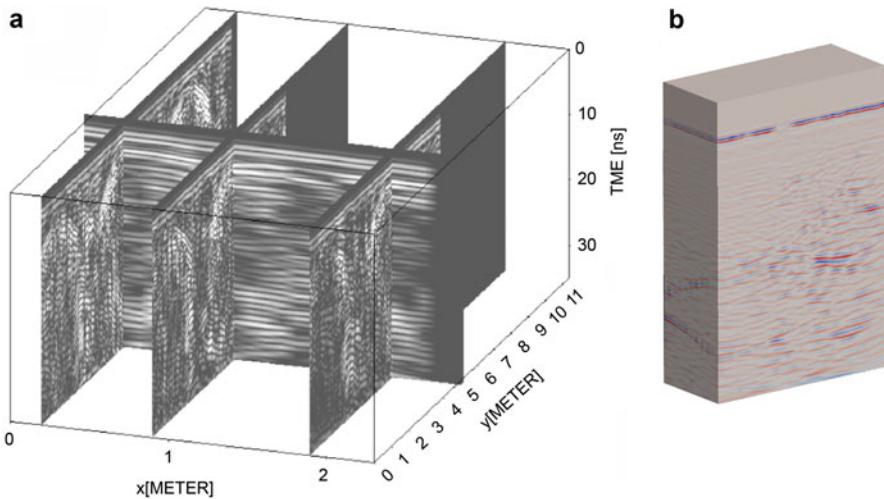
Figure 13 illustrates a C-scan. The A-scan results in one radar trace similar to any time-domain signal (Fig. 2). B-scans can be made by moving the GPR in one direction over the surface and making A-scans at predefined spatial steps, producing datagrams depicted as plane views. Finally, the recompilation of data over a surface is called a C-scan, which is presented and processed usually as slices of B-scan datagrams. At this point, it is worth remarking that it is possible to acquire data at different heights over the surface. Although this approach could be considered strictly as a C-scan, and it is useful to deal with first wave reflected from the host or the varying footprint of the antennas, the common PSTs are not focused on this method of collecting data. In fact, the convention is to speak of signal processing techniques when A-scans are under consideration, and image processing techniques

**Table 4** GPR software

Software	Application
GPR slice ( <a href="#">Goodman 2018</a> )	Archaeological and agriculture prospection
Radan ( <a href="#">GSSI 2018</a> )	Concrete, geology, forensics, and archaeological prospection
Radar studio ( <a href="#">USRadar 2018</a> )	Archaeological, humanitarian, and environmental prospection
Ibis Guardian ( <a href="#">GeoRadar 2018</a> )	Geology prospection
GPRMax ( <a href="#">GPRMax 2018</a> )	Electromagnetic simulation



**Fig. 12** A-, B-, and C-scans in a coordinate system



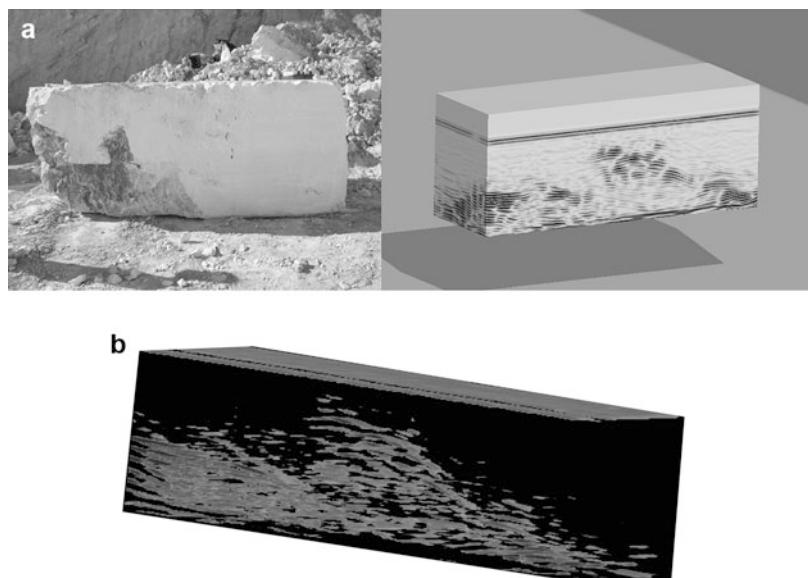
**Fig. 13** (a) Superposition of B-scans. (b) 3D interpolation. C-scan examples

when B-scans are treated as a unique group. Regardless of the use of one or more of these techniques, a subsequent stage called pattern recognition is aimed at predicting accurately the position and character of targets according to signal or image analysis or even soft-computing algorithms for learning and classifying purposes.

## Signal and Image Processing Techniques

The general objective of PSTs is to present simple information that can readily be interpreted by the operator. To this end, processing and prediction procedures transform, separate, and classify the detected electromagnetic signal properly. The first step in the treatment of acquired signals is to distinguish between reflections coming from different targets, discarding those coming from objects that are different from the actual target, and then mapping the relevant ones to form a new set of transformed signals or images ready to complete the prediction. Figure 14 shows an application of PSTs for the prospection for marble at 900 MHz and the predicted fractures.

In general, the unwanted reflections come from an early reflection by the air-structure interface or from late reflections from additional flaws in the host, which can appear as direct reflections from those objects or even as secondary reflections (those appearing due to refraction or scattering of waves). Reflections having the origin in the air-structure interface usually vary with position, due to the varying heights of measurement on the surface (a very common fact in open-air fields where vegetation or rocks influence the survey) or from unmatched operation of the antenna due to the coupling with the host medium. Regarding late reflections, only relevant signals (i.e., those with an appreciable intensity when compared to that from the real target) need to be taken into account. It is worth remarking that very different waveforms can appear depending of the characteristics of the reflection center (e.g., electrically small size objects vs. heterogenous layered media), which



**Fig. 14** (a) Marble and C-scan at 900 MHz. (b) Predicted fractures

challenges the PST ability to predict. In general, all the processes mentioned above can be called clutter signals.

Clutter is recognized as one of the major limitations of the GPR technique, because poor signal-to-clutter (SNC) ratios jeopardize the detection of the target. In some applications, for instance, if the target is buried at a shallow depth, the clutter is overspread on the targets response. In addition, incoherent random noise may appear in heterogeneous soils as well as disturbing in readings coming from electromagnetic interferences from external sources. Another relevant issue arises when clutter signals are also frequency-dependent, as is the case in lossy materials. Therefore, signal processing techniques primary goal is to remove the signals interfering with the targets response.

In general, the SNC and signal-to-noise (SNR) can be improved through simple strategies. For A-scans, the most basic procedures are the zero-offset removal and the noise and clutter reduction based either on time- or frequency-domain filtering. In some situations, a time-varying gain may improve the performance by enhancing late-time responses or offset removal. More elaborated strategies can be found in the literature, including deconvolution techniques, spectral analysis methods, or target resonances (Daniels 2004).

In this sense, the one-dimensional information of a A-scan waveform contains not only the response from buried targets but also undesired signals including the abovementioned first reflection from the surface, the receiver/transmitter EM interaction, thermal noise, interferences from other systems, and early reflections from shallow medium heterogeneity. In general, the amplitude of the target reflected field is small compared to the undesired signals, but it can be extracted as long as the target and clutter responses are separated in time. To this end, windowing filtering techniques can be applied. These can be considered as finite impulse response (FIR) filters, with a major advantage in terms of the linear-phase response which preserves the original waveform of the target reflected field. Time-gating is the simplest method used to eliminate all unwanted signals in a time frame. However, the choice of the starting and end time of the windowing is somewhat arbitrary, unless a priori physical parameters such as wave velocity propagation and depth of target are available.

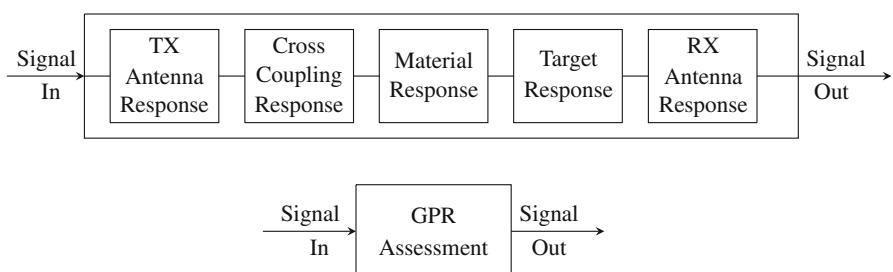
Since the EM wave propagation is attenuated with depth in the structure, most of the proprietary software provided by GPR manufacturers includes a time-variable gain processing. In this way, the gain can be adjusted with depth to amplify deeper reflections from possible targets, sometimes with customized gain functions. As a result, clutter and noise signals will be amplified as well. For this reason, this technique should be applied only when material properties are adequately assessed. Otherwise, the time-varying gain will add dispersive behavior to the actual waveform with unpredictable consequences over the success rate of the pattern recognition stage. There are several other techniques that can be applied in time domain including simple mean scan subtraction, background removal, and moving average filtering (Haykin and Van Veen 2007; Vetterli et al. 2014), but, as with the time-varying gain, their use can increase distortion and/or reduce SNC and SNR.

On the other hand, frequency-domain techniques offer an alternative outcome since they can select particular frequency sub-bands of interest in the GPR data. As

for time-domain techniques, there are many possible strategies (Haykin and Van Veen 2007; Madisetti 1997; Vetterli et al. 2014) for different purposes, commonly used to decrease some of the external interference signals or to remove the noise. In the first case, a high-pass filter also known by dewow (Daniels 2004) can be used to decouple the low-frequency interferences that can occur between transmitter and receiver antennas (Jol 2008). Also in this group, the noise appearing at intermediary frequency bands due to electromagnetic compatibility (EMC) issues can be extracted through notch filters (Vetterli et al. 2014), applying either finite (FIR) or infinite impulse response (IIR) filters. The latter has as a major drawback in the form of a nonlinear phase that distorts the waveform, and the former requires longer computational time for the application and therefore not practical for massive data.

For noise removal a low-pass filter can be applied, but this strategy usually subtracts the target response. The most preferred method is to consider the received signal as a convolution of several impulse responses. The GPR assessment can be considered a problem where the input signal is a designed waveform that will suffer modification as it passes through several blocks before returning to the receiver (Fig. 15). The initial signal is then convolved with the impulse responses of each block (e.g., the antenna or the host medium), considered as boundaries where reflection/transmission phenomena arise. Deconvolution techniques aim to separate the signals appearing at each block, and they can be performed either in spatial, time, or frequency domain. Deconvolution is a time-consuming task in the time domain, while in the frequency domain, it is performed by simple multiplications, but as long as the Fourier transform needs to be applied for the data, it cannot be foreseen a priori which method is advantageous. Another step to improve the performance of the deconvolution techniques is to consider a wavelet decomposition instead of Fourier transform, which opens new possibilities of filtering and amplifying waveforms aimed to increase SNC and SNR.

A final mention to complete this short review of A-scan techniques is made on the issue of discrimination of targets through target resonances and also to name other successful spectral-analysis strategies. In the first case, the underlying idea is that every object possess unique resonant features calculated by Prony's method (Blaricum and Mittra 1975). Through the knowledge of amplitude and phases of the resonances, more efficient filters and/or amplifiers can be applied thus enhancing



**Fig. 15** GPR transfer function

the performance of PSTs. Spectral-analysis methods, such as autoregressive moving average estimation or multiple signal classification, are usually applied in FMCW radars, where measured data are acquired directly in the frequency domain.

B-scan signal processing techniques take advantage of the higher information achieved by a spatial aggregation of measurements. In this sense, it is important to realize that a common weak point of A-scan procedures is that they need to include a high response from the target, which usually happens when the GPR equipment is on top of the target. Obviously, in most of the open scenarios, this location remains unknown, and the first issue is how to detect possible location of possible targets. To address this problem, the migration method is a deconvolution technique that allows refocusing of detected electromagnetic fields to their true spatial and temporal location (Leuschen and Plumb 2001). Alternative approaches based in time and frequency domain are, respectively, Kirchhoff and frequency-wavenumber migration techniques (Liu et al. 2017; Sakamoto et al. 2015). On the other hand, the synthetic-aperture processing combines the responses from the directive beams of antennas in B-scan to produce holographic SAR images (Skolnik 1980). Finally, B-scan information can be treated through image processing techniques (González-Huici 2013), which have been proven to be effective in detecting edges of objects and in improving visual inspection through image filtering.

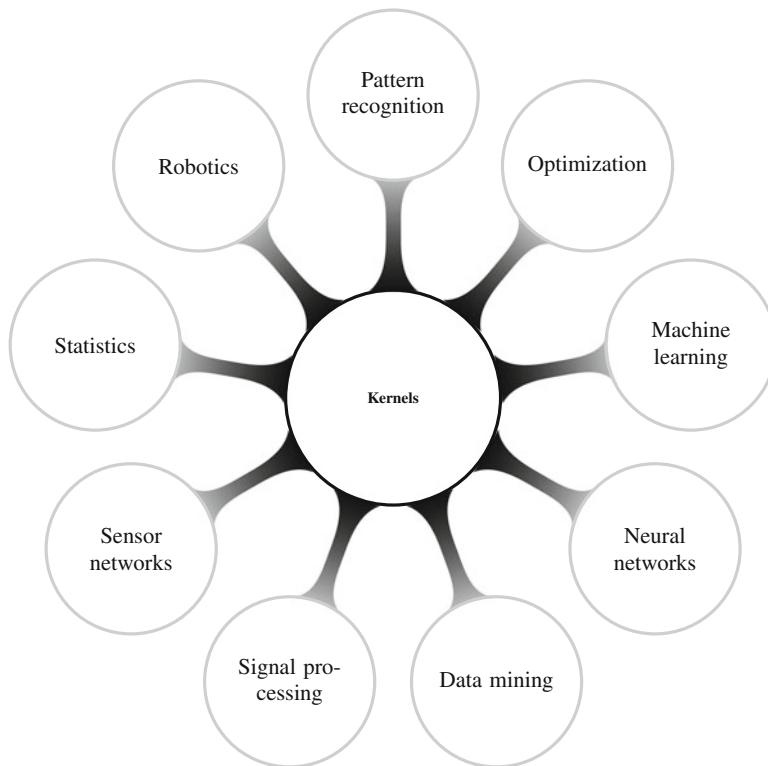
There is no technique that can work for all possible GPR configurations and applications. Some of them are not suitable for the GPR data available, or simply the resources needed to be implemented cannot be afforded, or that the goals of the PST includes too many challenges. As an example of the latter case, the use of GPR for landmine elimination requires specific techniques aimed not only at the location of the target but also at providing sufficient information for safe extraction (Gader et al. 1999; Nicolaescu 2013; Missaoui et al. 2011).

## Pattern Recognition

In GPR pattern recognition, a controlled training set of waveforms recorded from the receiving antenna is provided for several training scenarios. With this information, a mathematical or a theoretical procedure, usually involving soft-computing techniques, is derived as a predictive system. The quality of the predictive strategy can be further assessed for any attempt to predict the sources that created a new set of reflections coming from targets located in GPR scenarios not contained in the training set. The prediction of scattering sources through electromagnetic waves is a well-known ill-posed inverse problem, as opposed to the well-posed direct problems aimed at calculating the electromagnetic fields in physical scenarios where geometrical parameters or material properties are known. For NDT applications using GPR, pattern recognition aims at determining a useful finite number of parameters which characterize targets in a given medium by means of identification of their electrical and geometrical properties. In general, this is a challenging problem because sometimes the available data is not sufficient for the prediction, which usually can be addressed thorough the insertion of some additional

information of the GPR scenario, such as symmetries or other simplifying assumptions. On the other hand, in other cases a large number of data are available, but, as long as noise and other external sources can be present in the data, the information contains some uncertainties which can bring the prediction system to unstable or inaccurate solutions.

The purpose of pattern recognition techniques is to improve the GPR assessment and diminish the importance of human interpretation by solving the inverse problem. In general, they are applied after signal processing techniques have been applied. To be effective, pattern recognition techniques should be cost-effective and adaptable to different applications, give a fast response for big amount of data, and provide a desirable false alarm response or probability of detection. Considering the purpose of using GPR for NDT, we assume that it is impossible to develop a technique that can classify and distinguish buried targets in all types of GPR assessments. There are several applications using different kernel-based methods. According to Scholkopf and Smola (2001), a kernel function  $k(x, x')$  characterizes the similarity between classes of objects. In the GPR context, the objects are the received samples (time-domain) or patterns in the image (space-domain). Kernels are used in several applications as sketched in Fig. 16.



**Fig. 16** Kernel-based modeling

Kernel-based models have the advantage of being free of local minima and the sparseness of the solution (Shawe-Taylor and Cristianini 2004). On the other hand, the choice of a kernel for a given application is a difficult task. There are several types of kernels available. Some of them are shown in Table 5. Kernel-based methods also suffer from computational burden when implemented. That is due to the fact that Kernels operate on GPR data as though it were projected into a higher dimensional space.

In general, different techniques can be applied to GPR data, including, but are not limited to, principal component analysis, discriminant analysis, Markov models, decision trees, k-nearest neighbors, edge histogram descriptors, spectral features, Bayesian classifiers, geometric features, texture features, neural networks, and support vector machines. Some of these are discussed next.

### Principal Component Analysis

In some situations, the dimension of the input vector is large, but the components of the vectors are highly correlated (redundant). Principal component analysis (PCA) is a way of identifying patterns in data and expressing the data in such a way as to highlight their similarities and differences (Rodriguez et al. 2015). The main advantage of PCA is that once you have found these patterns in the data, the data may be compressed, i.e., by reducing the number of dimensions, without much loss of information. The technique is mostly used in image compression.

This technique has three effects: it orthogonalizes the components of the input vectors (so that they are uncorrelated with each other), it orders the resulting orthogonal components (principal components) so that those with the largest variation come first, and it eliminates those components that contribute the least to the variation in the data set. The input vectors are first normalized so that they have zero mean and unity variance. For PCA to work properly, one has to subtract the mean from each of the data dimensions.

The PCA uses linear mapping of a given set of samples  $S_q = \{x_1, \dots, x_T | x_i \in \mathbb{R}^p\}$  to construct a new data set  $S_p = \{y_1, \dots, y_T | y_i \in \mathbb{R}^q\}$ , where  $q \leq p$ . Considering a  $p \times q$   $V_q$  matrix the PCA algorithm can be described as follows:

1. Subtract the mean from each data dimension.
2. Calculate the covariance matrix.
3. Calculate the eigenvectors and eigenvalues of the covariance matrix.

**Table 5** Examples of Kernels (Scholkopf and Smola 2001)

Type of Kernel	Function	Comments
Polynomial	$k(x, x') = \langle x, x' \rangle^d$	Homogeneous polynomial
Gaussian	$k(x, x') = \exp\left(\frac{\ x-x'\ ^2}{2\sigma^2}\right)$	Gaussian radial basis
Sigmoid	$k(x, x') = \tanh(\kappa \langle x, x' \rangle + \theta)$	Gaussian with $\sigma > 0$
Triangular (Fleuret and Sahbi 2003)	$k(x, x') = 1 - \frac{\ x-y\ }{\sigma}$	Eucledian distance

4. Choose components and form a feature vector.
5. Derive the new data set.

Another interpretation of the PCA is the identification of directions that maximize the variance. The transformation  $V_q$  generates a projection space in which the covariance matrix is diagonal. The diagonal covariance matrix implies that the variance of a variable with itself is maximized but is minimized with any other variable. Thus, the  $q$  variables with higher variance in the new space should be kept. The principal components of a set of data in  $\mathbb{R}^p$  provide a sequence of best linear approximations to that data, of all ranks  $q \leq p$ .

## Discriminant Analysis

Discriminant analysis is a pattern recognition technique which classifies the data using a better mapping than PCA. In general, it is applied after PCA to parameterize decision boundaries in a more effective way since it calculates the best discriminating components of the sample data. However, there is a trade-off to be considered in defining the PCA dimensional space prior to discriminant analysis in order to avoid the curse of dimensionality, an ill-conditioned problem (Sun and Li 2003), and a rise in the computational effort. However this method usually fails for nonlinear GPR problems.

## Feature Selection

Feature selection is a process commonly used in machine learning, wherein a subset of the features available from the data is selected for application of the learning algorithm. It is different from feature extraction because the latter creates new features based on some combination of the original ones (Dash and Liu 2007). The best subset contains the least number of dimensions that most contribute to accuracy. The remaining unimportant dimensions are discarded. This is an important stage of preprocessing aimed to avoiding the well-known curse of dimensionality (another way is feature extraction) and, therefore, guarantee adequate convergence of the learning algorithm. It can also provide some understanding concerning the nature of the problem, as it indicates the main physical properties needed to classify an underground target. Therefore, feature selection is the task of choosing a small subset of features which is sufficient to predict the target and can capture the relevant information about the GPR problem.

The existing literature distinguishes between two types of feature selection algorithms: the so-called wrapper and filter methods (Jain and Zongker 1997). Wrapper methods, which are computationally intensive and tend to over-fit (Křížek et al. 2007), estimate the usefulness of a subset of features by a given predictor or learning algorithm. The wrapper methods try to directly optimize the performance of a given predictor. This is done by estimating the generalization performance. On the other hand, filter or variable ranking methods compute relevance scores for each single feature and choose the most relevant ones according to those scores. This can be done using, usually in an ad hoc manner, evaluation functions aiming at searching the set of features that maximize the information. Among the commonly used

evaluation functions are the mutual information, the margin, and the dependence measures. The main drawback of such simple filtering methods is that they are not able to detect inter-feature dependencies. One example of these latter dependencies in the GPR problem is the relation between clutter and antenna interference that can occur at the same time.

Feature selection commonly found in GPR applications are spectral features, geometric features, and text features. They also can be used in conjunction with other techniques such as neural networks or support vector machines (Kwak and Choi 2002; Nguyen and de la Torre 2010).

### **Markov Models**

Markov models have been widely used in different fields of applications including speech and handwriting recognition or to automatically align multiple biological sequences. The principle is based on modeling the statistical behavior of a time-varying parameter. Considering GPR, it can be used to automatically divide the received waveform into multiple sequence patterns. Those patterns can then be related to a database of possible targets by its signature (Gader et al. 2001; Manandhar et al. 2015).

A hidden Markov model (HMM) consists of states and edges connecting those states. At each recorded sample time (or observation time), the process may be in one of these states. Since the actual state is not observable, it is called hidden. In this structure there are transition probabilities between states that are connected by an edge. The HMM will be trained to establish the likelihood of each sample to be correlated to an objective. In this sense, the GPR problem works in a similar fashion since data is recorded in time intervals over the structure under test. In addition, for this application, the states in HMM are finite. Furthermore, an HMM is called continuous if the state probability density functions are continuous and discrete if the state probability density functions are discrete (Gader et al. 1999). HMM has proven to be a valuable tool for pattern recognition. However, a balance between important features in the scattered wave should be sought in order to improve the HMM performance (Missaoui et al. 2011).

---

### **Summary**

This chapter has briefly discussed the use GPR as an NDT tool. In contrast to other NDT methods, GPR is based on EM wave propagation which provides advantages and limitations. Among the advantages, we can cite relatively low cost of survey, compact size of equipment, excellent trade-off between resolution and depth of detection, and suitability to an overwhelming variety of applications. However, the analysis the data is not straightforward not only by the complexity of the physical phenomena involved in propagation and scattering of electromagnetic waves but also for the impact of environmental conditions on the electromagnetic description of the media and consequently on the data acquired. To reduce this burden, the GPR equipment allows multiple configurations which have to be carefully explored and

selected in the planning of the survey. Even with the right choices of those configurations, the ability to predict and detect targets requires considerable expertise as well as a set of specialized postprocessing tools. Far from being a closed line of research, the possibilities to improve detection are still open, and more exciting contributions are expected in the future to create a more powerful NDT tool.

---

## References

- Alkhalifeh K et al (2016) Efficient mom simulation of 3-d antennas in the vicinity of the ground. *IEEE Trans Antennas Propag* 64(12):5335–5344
- Almeida ER et al (2016) Microwave tomography-enhanced gpr in forensic surveys: the case study of a tropical environment. *IEEE J Sel Top Appl Earth Obs Remote Sens* 9(1):115–124. <https://doi.org/10.1109/JSTARS.2015.2466556>
- Al-Qadi IL, Hazim OA, Su W, Riad SM (1995) Dielectric properties of Portland cement concrete at low radio frequencies. *J Mater Civ Eng* 7(3):192–198. [https://doi.org/10.1061/\(ASCE\)0899-1561\(1995\)7:3\(192\)](https://doi.org/10.1061/(ASCE)0899-1561(1995)7:3(192))
- Balanis C (2005) Antenna theory: analysis and design, vol 1. Wiley, Hoboken
- Blaricum MV, Mittra R (1975) A technique for extracting the poles and residues of a system directly from its transient response. *IEEE Trans Antennas Propag* 23(6):777–781. <https://doi.org/10.1109/TAP.1975.1141184>
- Bykztrk O, Yu T, Alberto Ortega J (2006) A methodology for determining complex permittivity of construction materials based on transmission-only coherent, wide- bandwidth free-space measurements. *Cem Concr Compos* 28:349–359
- Caorsi S, Cevini G (2005) An electromagnetic approach based on neural networks for the gpr investigation of buried cylinders. *IEEE Geosci Remote Sens Lett* 2(1):3–7. <https://doi.org/10.1109/LGRS.2004.839648>
- Cook JC (1975) Radar transparencies of mine and tunnel rocks. *Geophysics* 40(5):865–885. <https://doi.org/10.1190/1.1440573>
- Daniels DJ (2004) Ground penetrating radar, 2nd ed. IEE London, UK
- Dash M, Liu H (2007) Dimensionality Reduction. Wiley. <https://doi.org/10.1002/9780470050118.ece112>
- De Coster A et al (2016) Fundamental analyses on layered media reconstruction using gpr and full-wave inversion in near-field conditions. *IEEE Trans Geosci Remote Sens* 54(9):5143–5158. <https://doi.org/10.1109/TGRS.2016.2556862>
- Donato LD, Crocco L (2015) Model-based quantitative cross-borehole gpr imaging via virtual experiments. *IEEE Trans Geosci Remote Sens* 53(8):4178–4185. <https://doi.org/10.1109/TGRS.2015.2392558>
- Du et al (2016) Research of fusion method of gpr tomography images based on wavelet transform. In: 2016 16th international conference on ground penetrating radar (GPR), Piscataway, pp 1–6. <https://doi.org/10.1109/ICGPR.2016.7572666>
- Ebihara S et al (2012) Interference criterion for coaxial-fed circular dipole array antenna in a borehole. *IEEE Trans Geosci Remote Sens* 50(9):3510–3526. <https://doi.org/10.1109/TGRS.2011.2182517>
- Evans S (1963) Radio techniques for the measurement of ice thickness. *Polar Rec* 11(73):406410. <https://doi.org/10.1017/S0032247400053523>
- Fleuret F, Sahbi H (2003) Scale-invariance of support vector machines based on the triangular kernel. In: 3rd international workshop on statistical and computational theories of vision
- Gader et al (1999) Applications of hidden Markov models to detecting land mines with ground-penetrating radar. <https://doi.org/10.1117/12.356989>
- Gader PD et al (2001) Landmine detection with ground penetrating radar using hidden Markov models. *IEEE Trans Geosci Remote Sens* 39(6):1231–1244. <https://doi.org/10.1109/36.927446>

- GeoRadar (2018) Ibis guardian. <https://idsgeoradar.com/products/software/ibisguardian>. Accessed: 2018-02-12
- Giannakis I et al (2016) A realistic fddt numerical modeling framework of ground penetrating radar for landmine detection. *IEEE J Sel Top Appl Earth Obs Remote Sens* 9(1): 37–51. <https://doi.org/10.1109/JSTARS.2015.2468597>
- Gilad-Bachrach R, Navot A, Tishby N (2004) Margin based feature selection – theory and algorithms. In: Proceedings of the twenty-first International conference on machine learning. ACM, New York, ICML '04, p 43
- González-Huici M (2013) Accurate ground penetrating radar numerical modeling for automatic detection and recognition of antipersonnel landmines
- Goodman (2018) Gpr slice. <http://www.gpr-survey.com>. Accessed: 2018-02-12
- GPRMax (2018) Gprmax. <http://www.gprmax.com>. Accessed: 2018-02-12
- GSSI (2018) Radan software. <https://www.geophysical.com/software>. Accessed: 2018-02-12
- Guan B et al (2017) Near-field full-waveform inversion of ground-penetrating radar data to monitor the water front in limestone. *IEEE J Sel Top Appl Earth Obs Remote Sens* 10(10):4328–4336. <https://doi.org/10.1109/JSTARS.2017.2743215>
- Haykin S, Van Veen B (2007) Signal and systems, 2nd edn. Wiley India Pvt. Limited
- Hfinghoff JF, Overmeyer L (2013) Resistive loaded antenna for ground penetrating radar inside a bottom hole assembly. *IEEE Trans Antennas Propag* 61(12):6201–6205. <https://doi.org/10.1109/TAP.2013.2283604>
- Hulsenbeck et al (1926) German patent. No. 489434
- Ida N (2007) Engineering electromagnetics
- Jain A, Zongker D (1997) Feature selection: evaluation, application, and small sample performance. *IEEE Trans Pattern Anal Mach Intell* 19(2):153–158. <https://doi.org/10.1109/34.574797>
- Jol H (2008) Ground penetrating radar theory and applications. Elsevier Science, Amsterdam
- Čížek P et al (2007) Improving stability of feature selection methods. Springer, Berlin/Heidelberg, pp 929–936
- Kwak N, Choi CH (2002) Input feature selection for classification problems. *IEEE Trans Neural Netw* 13(1):143–159. <https://doi.org/10.1109/72.977291>
- Leckebusch J (2011) Comparison of a stepped-frequency continuous wave and a pulsed gpr system. *Archaeol Prospect* 18(1):15–25. <https://doi.org/10.1002/arp.396>
- Lestari AA et al (2010) A modified bow-tie antenna for improved pulse radiation. *IEEE Trans Antennas Propag* 58(7):2184–2192. <https://doi.org/10.1109/TAP.2010.2048853>
- Leuschen CJ, Plumb RG (2001) A matched-filter-based reverse-time migration algorithm for ground-penetrating radar data. *IEEE Trans Geosci Remote Sens* 39(5):929–936
- Liu H et al (2015) A hybrid dual-polarization gpr system for detection of linear objects. *IEEE Antennas Wirel Propag Lett* 14:317–320. <https://doi.org/10.1109/LAWP.2014.2363826>
- Liu X et al (2017) Ground penetrating radar data imaging via Kirchhoff migration method. In: 2017 international applied computational electromagnetics society symposium Italy (ACES), pp 1–2
- Madisetti V (1997) The digital signal processing handbook. Electrical engineering handbook. Taylor & Francis
- Manandhar A et al (2015) Multiple-instance hidden Markov model for gpr-based landmine detection. *IEEE Trans Geosci Remote Sens* 53(4):1737–1745. <https://doi.org/10.1109/TGRS.2014.2346954>
- Mehta PK (1986) Concrete: structure, properties and materials. Prentice Hall, Englewood Cliffs
- Missaoui O et al (2011) Land-mine detection with ground-penetrating radar using multistream discrete hidden Markov models. *IEEE Trans Geosci Remote Sens* 49(6): 2080–2099. <https://doi.org/10.1109/TGRS.2010.2090886>
- Neal A (2004) Ground-penetrating radar and its use in sedimentology: principles, problems and progress. *Earth Sci Rev* 66(3):261–330. <https://doi.org/10.1016/j.earscirev.2004.01.004>
- Nguyen MH, de la Torre F (2010) Optimal feature selection for support vector machines. *Pattern Recogn* 43(3):584–591. <https://doi.org/10.1016/j.patcog.2009.09.003>

- Nicolaescu I (2013) Improvement of stepped-frequency continuous wave ground- penetrating radar cross-range resolution. *IEEE Trans Geosci Remote Sens* 51(1):85–92
- Pieraccini M (2018) Noise performance comparison between continuous wave and stroboscopic pulse ground penetrating radar. *IEEE Geosci Remote Sens Lett* 15(2):222–226. <https://doi.org/10.1109/LGRS.2017.2781458>
- Qin H, Xie X (2016) Design and test of an improved dipole antenna for detecting enclosure structure defects by cross-hole gpr. *IEEE J Sel Top Appl Earth Obs Remote Sens* 9(1): 108–114. <https://doi.org/10.1109/JSTARS.2015.2466450>
- Queiroz FAA, Vieira DAG, Travassos XL (2013) Analyzing the relevant features of gpr scattered waves in time- and frequency-domain. *Res Nondestruct Eval* 24(2):105–123. <https://doi.org/10.1080/09349847.2012.752889>
- Raimundo et al (2014) Frequency modulated interrupted continuous wave signals in different radar imaging applications. In: 2014 XXXIth URSI general assembly and scientific symposium (URSI GASS), pp 1–4. <https://doi.org/10.1109/URSIGASS.2014.6929601>
- Reeves B (2014) Noise modulated gpr: second generation technology. In: Proceedings of the 15th international conference on ground penetrating radar, pp 708–713. <https://doi.org/10.1109/ICGPR.2014.6970519>
- Rodriguez JB et al (2015) A prediction algorithm for data analysis in gpr-based surveys. *Neurocomputing* 168:464–474. <https://doi.org/10.1016/j.neucom.2015.05.081>
- Sagnard F, Rejiba F (2011) Wide band coplanar waveguide-fed bowtie slot antenna for a large range of ground penetrating radar applications. *IET Microwaves, Antennas Propag* 5(6):734–739
- Sakamoto et al (2015) Frequency-domain Kirchhoff migration for near-field radar imaging. In: 2015 I.E. conference on antenna measurements applications (CAMA), pp 1–4
- Scholkopf B, Smola AJ (2001) Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT Press, Cambridge, MA
- Shao J et al (2014) Tem horn antenna loaded with absorbing material for gpr applications. *IEEE Antennas Wirel Propag Lett* 13:523–527. <https://doi.org/10.1109/LAWP.2014.2311436>
- Shawe-Taylor J, Cristianini N (2004) Kernel methods for pattern analysis. Cambridge University Press, New York
- Skolnik M (1980) Introduction to radar systems. Electrical engineering series. McGraw-Hill, New York
- Sun Y, Li J (2003) Time-frequency analysis for plastic landmine detection via forward-looking ground penetrating radar. *IEE Proc Radar Sonar Navig* 150:253
- USRadar (2018) Radar studio. <http://www.usradar.com/ground-penetrating-radar-gpr-software/radar-studio/>. Accessed: 2018-02-12
- Vetterli M et al (2014) Foundations of signal processing. Cambridge University Press, Cambridge
- Yang CC, Bose NK (2005) Landmine detection and classification with complex-valued hybrid neural network using scattering parameters data set. *IEEE Trans Neural Netw* 16(3):743–753. <https://doi.org/10.1109/TNN.2005.844906>
- Zeng Z et al (2015) Improving target detection accuracy based on multipolarization mimo gpr. *IEEE Trans Geosci Remote Sens* 53(1):15–24. <https://doi.org/10.1109/TGRS.2014.2312937>

---

## **Part V**

### **X-Ray Techniques**



# Processing of X-Ray Images

29

Sergei Chakhlov

## Contents

Introduction .....	1028
General Procedure .....	1028
Digitizing and Image Formats .....	1028
Image Histogram .....	1029
Lookup-Table and Contrast Enhancement .....	1029
Background Removal .....	1031
Frame Averaging .....	1033
Offset, Gain, and Reference Channel Calibrations .....	1033
Geometric Transformations .....	1036
Image Stitching .....	1037
Image Filtering .....	1038
Edge Detection .....	1040
Segmentation .....	1042
Computer Vision Problems (Features Extraction and Selection for Classification) .....	1045
Dual Energy Imaging .....	1047
Stereoscopic 3D Images .....	1050
Summary .....	1051
References .....	1051

## Abstract

A general overview of X-ray image processing is presented. In this chapter, the simplest and most effective digital image processing algorithms and visualization techniques are briefly considered. The algorithms and techniques are illustrated by actual X-ray images. The most developing areas of X-ray image processing are outlined. References to X-ray image database, formats, and software are given.

---

S. Chakhlov (✉)

School of Non-Destructive Testing and Security, Tomsk Polytechnic University, Tomsk, Russia  
e-mail: [chakhlov@tpu.ru](mailto:chakhlov@tpu.ru)

## Introduction

Digital image processing (Pratt 2007; Gonzalez and Woods 2008; Castleman 1996) is the use of computer algorithms to perform image processing on digital images. As a subcategory or the field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the buildup of noise and signal distortion during processing.

---

## General Procedure

The general steps of X-ray image processing are:

- Digitizing and saving the image
- Image enhancement, e.g., histogram and lookup tables
- Removing of bad pixels and background
- Offset, gain, and reference channel calibrations
- Geometric corrections
- Image filter techniques (averaging, median, unsharp mask, bilateral filters)
- Edge detection (Roberts, Sobel, Canny, emboss filters)
- Segmentation
- Classification
- Dual energy imaging
- Anaglyph 3D images

---

## Digitizing and Image Formats

There are three X-ray image sources:

1. X-ray film scanners
2. X-ray sensitive plates
3. Linear and matrix detector arrays

Each type of these sources in combination with X-ray radiation generators has its own specific distortions (Mery 2015). Therefore, this chapter will be focused on the widely used and the most effective X-ray image processing techniques and some information on the visualization of X-ray images will be given.

Currently, the most popular X-ray image formats are the following: TIFF, PNG, and DICONDE. In these formats, data compression without loss of quality is mainly used. The DICONDE format (Jobst et al. 2010) is actively promoted by the General Electric multinational corporation and is enforced by the standard ASTM 2339 (ASTM 2008). The halftone images are most commonly used with a 16-bit per

pixel (16-bpp) resolution and a black-and-white color palette. Pseudocolors are not practically used for X-ray images, excluding the dual energy case (Mery 2015).

Examples of X-ray images can be found in a large public collection of images (Mery et al. 2015) (mainly with 8-bits per pixel).

There are many software programs to visualize and process images but for the X-ray images we recommend (ISee! 2017; Mery 2018).

Currently, the most actively developing areas of the X-ray image processing are: automatic pattern recognition (for medicine, baggage, and cargo screening), 3D recognition, computed tomography, reverse engineering, etc.

## Image Histogram

An image histogram acts as a graphical representation of the tonal distribution in a digital image. It plots the number of pixels for each tonal value. By looking at the histogram for a specific image, a viewer will be able to judge the entire tonal distribution at a glance.

## Lookup-Table and Contrast Enhancement

The human eye is only capable to resolve around 40 gray levels (Castleman 1996); however, for computer vision applications, gray-level resolution must be a minimum of 256 levels. Therefore, a lookup table (LUT) is used to transform the input 16-bpp pixel lightness values ( $2^{16} = 65,536$ ) into the 8-bpp output computer display format.

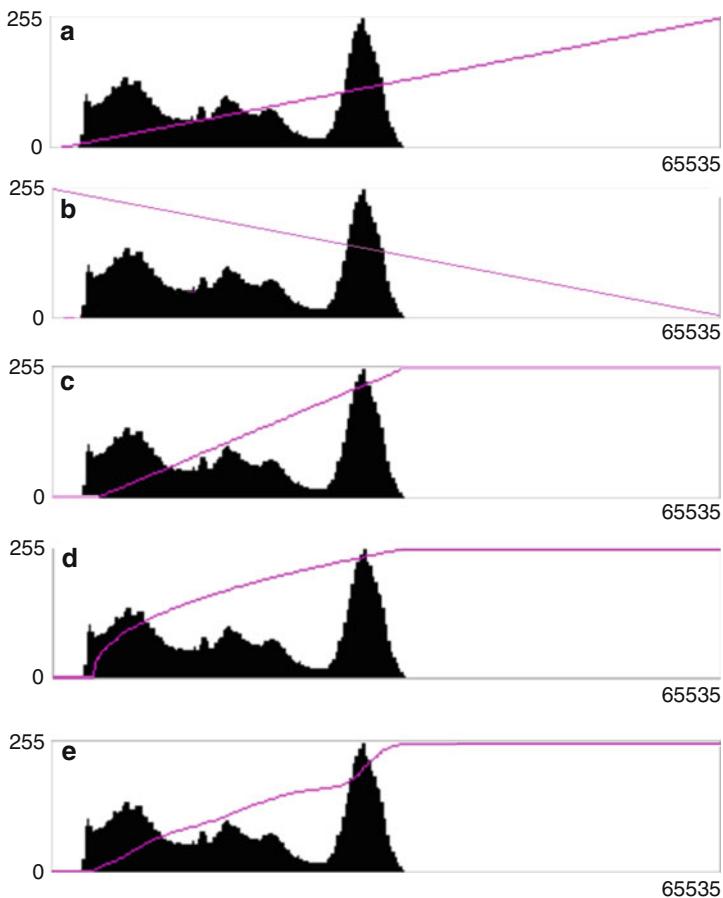
Low contrast is the most common defect due to the limited range of reproducible lightness (Fig. 1a). The lightness range of the output image can be increased by converting the value of the output signal  $T$  depending only on the input value  $t$ :

$$T = f(t). \quad (1)$$

The LUT function  $f(x)$  must be nonnegative and not a decreasing one. Figure 1 shows examples of different helpful LUTs (magenta curves): a – standard LUT, b – negative, c – piecewise linear LUT, d – piecewise power LUT ( $T = t^{\gamma}$ ,  $\gamma = 0.45$ ), e – global histogram equalization LUT, obtained by integrating the image histogram (shown by a black color):

$$y(x) = \int_0^x hist(t)dt. \quad (2)$$

Figure 2 shows the images, corresponding to these LUTs. Figure 2e is similar to Fig. 2c excluding more contrast bright areas. It is known that the subjective

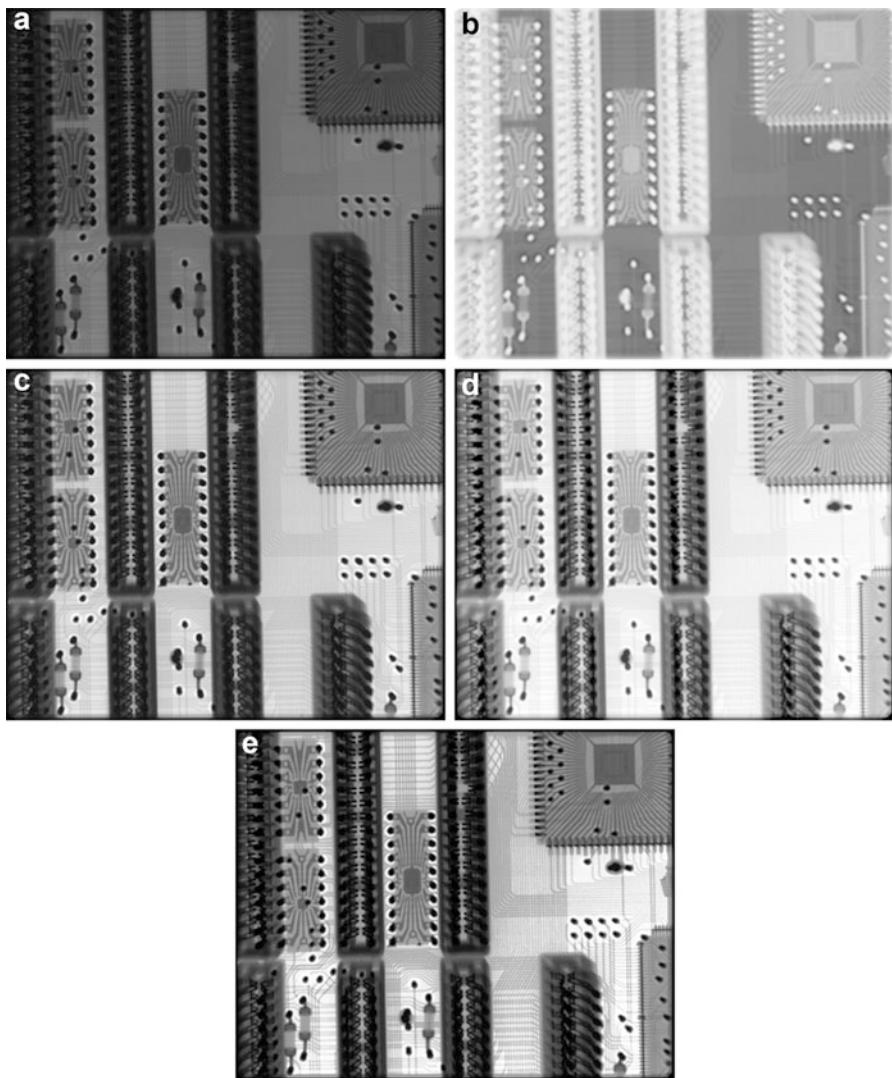


**Fig. 1** Histograms and LUTs for contrast enhancement: (a) original image, (b) negative, (c) linear transform, (d) power, (e) global histogram equalization

brightness is a power function ( $\gamma = 0.33$ ) of luminance (Hanes 1949) because such a type of LUT is very helpful to show dark images.

Ordinary histogram equalization uses the same transformation derived from the image histogram to transform all pixels. This works well when the distribution of pixel values is similar throughout the image. However, when the image contains regions that are significantly lighter or darker than most of the image, the contrast in those regions will not be sufficiently enhanced.

Contrast limited adaptive histogram equalization (CLAHE) method was developed to overcome this shortcoming. CLAHE (Zuiderveld 1994) computes several histograms, each corresponding to a distinct section of the image, and uses them to redistribute the lightness values of the image by bilinear interpolation. Figure 3 shows the image transformed by CLAHE and its resulting histogram (CLAHE divides the image into  $4 \times 4 = 16$  parts and increases contrast by a factor of 8).



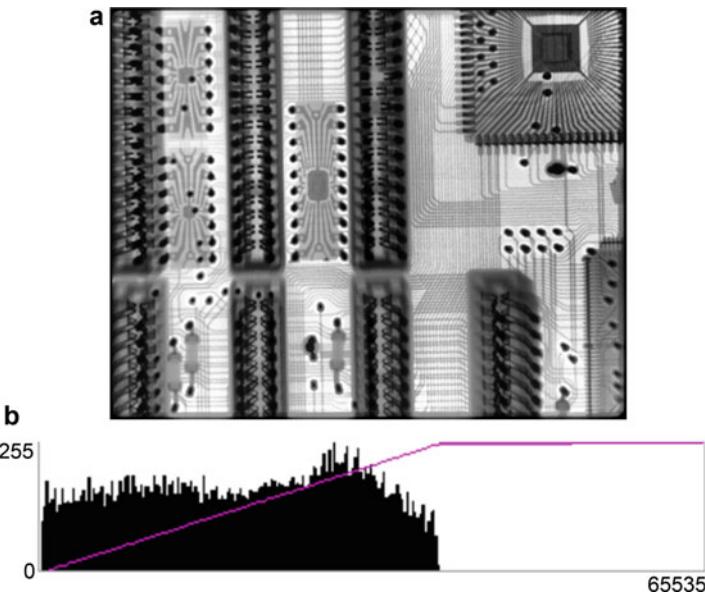
**Fig. 2** Examples of contrast enhancement: (a) original image, (b) negative, (c) linear transform, (d) power transform, (e) histogram equalization

---

## Background Removal

The difference between the images  $f(x,y)$  and  $b(x,y)$  is expressed by the formula:

$$g(x,y) = f(x,y) - b(x,y) + \text{const.} \quad (3)$$



**Fig. 3** (a) Image transformed by CLAHE, (b) its histogram

If it is possible to acquire the X-ray image of some object and the background image without this object, then a difference of the two images gives a clear image of the object only (see Fig. 4).

When the background image is absent, one can estimate it by the first order least square approximation:

$$b(x,y) \approx A_0 + A_1x + A_2y, \quad (4)$$

or second order approximation:

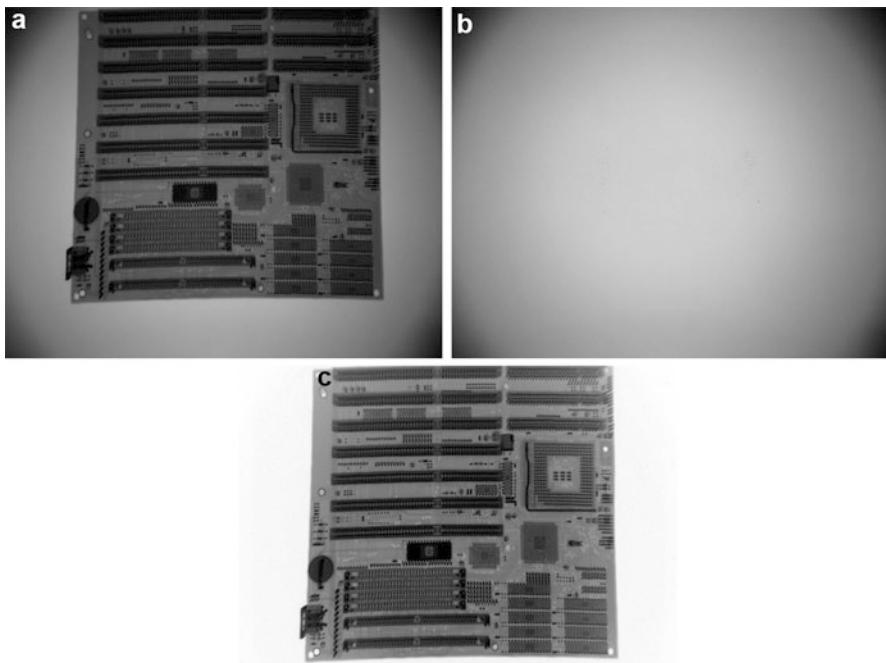
$$b(x,y) \approx B_0 + B_1x + B_2y + B_3x^2 + B_4xy + B_5y^2 \quad (5)$$

of the original image with the object.

Another way is an estimation of each background pixel by averaging over the large neighborhood  $(2m+1) \times (2n+1)$  of the pixel  $f_{ij}$ :

$$b_{ij} = \frac{1}{(2m+1) \times (2n+1)} \sum_{p=-m}^m \sum_{q=-n}^n f_{i+p,j+q}. \quad (6)$$

Figure 5 shows the usage of the background image estimated by the averaging over  $31 \times 31$  pixels. The disadvantage of this approach is an excessive emphasis of the gray gradients due to the wrong estimation of background for the gradient regions. A large-scale averaging filter can be calculated fastly by an algorithm from (Rakshit et al. 2007).



**Fig. 4** Background removal: (a) X-ray image of object, (b) the image without object, (c) difference of (a) and (b) images

---

## Frame Averaging

Averaging is used to remove X-ray quantum noise. This technique requires  $n$  stationary X-ray images. It computes the averaged image as follows:

$$F(x,y) = \frac{1}{n} \sum_{k=1}^n f_k(x,y) \quad (7)$$

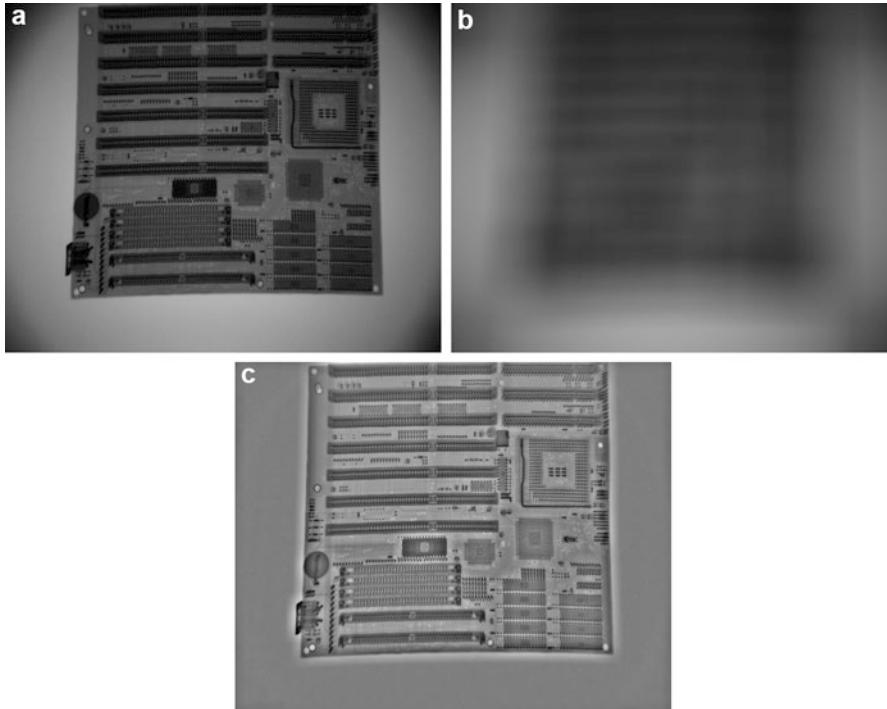
This operation improves the signal-to-noise ratio by a factor of  $\sqrt{n}$  (Castleman 1996).

---

## Offset, Gain, and Reference Channel Calibrations

Such calibration is usually used to calibrate images obtained with a detector array, since each detector has its own individual characteristics:

$$f_k^{\text{new}} = \frac{(f_k^{\text{old}} - B_k)}{(W_k - B_k)} \times 60000, \quad (8)$$

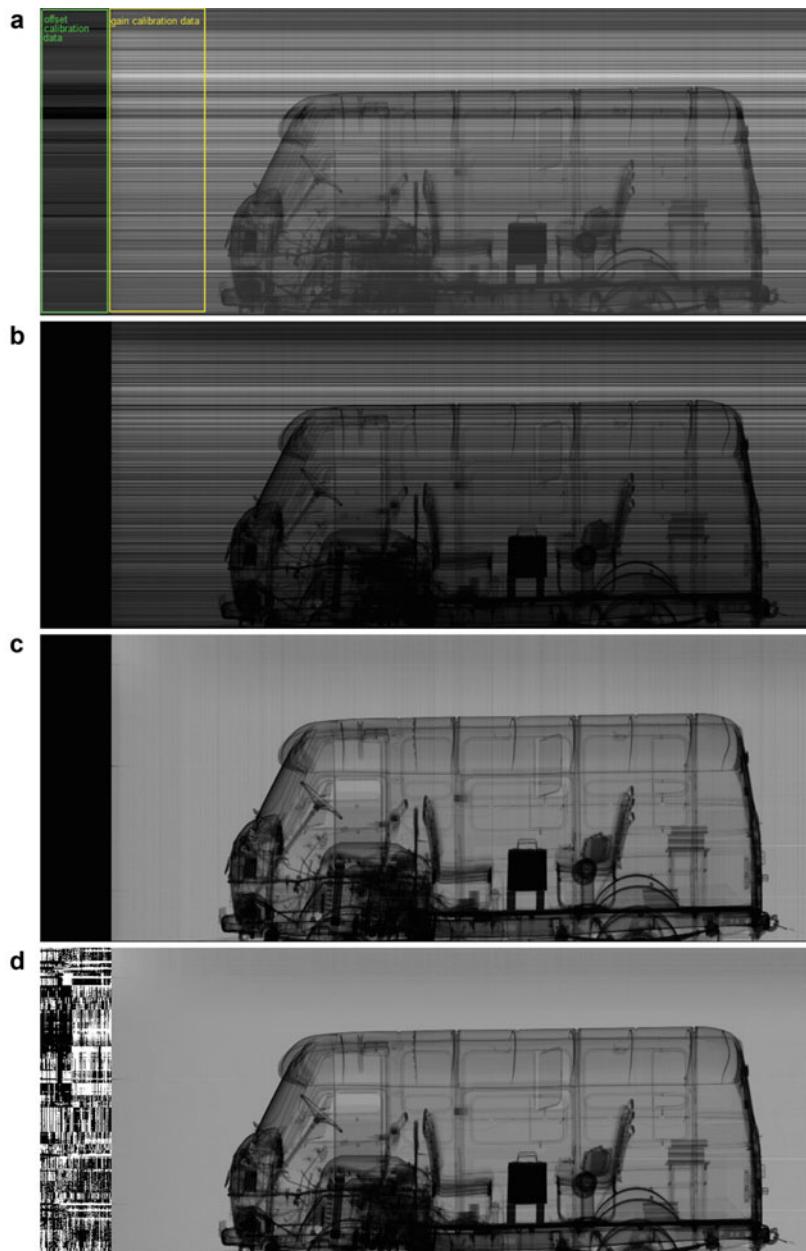


**Fig. 5** Removal of background estimated by the averaged values: (a) original image, (b) image averaged by mask  $31 \times 31$  (background estimation), (c) difference of (a) and (b) images

where  $f_k^{\text{new}}$  is the calibrated value,  $f_k^{\text{old}}$  is the original value,  $B_k$  is the offset value (without X-ray),  $W_k$  is the gain value (with X-ray and without inspected object). The  $k$  index is a detector number. The scale factor of 60,000 is selected, because otherwise the calibrated value may exceed the maximal possible pixel value of 65,535. The offset and gain values are usually calculated by averaging the values of the detectors without X-ray and with X-ray but without the inspected object, correspondingly. The offset values were calculated by the green outlined part of the initial image, produced without X-ray Fig. 6a. The gain values were calculated by the yellow outlined part of the image without the inspected object.

Figure 6a, b, c show the initial, offset, and gain calibrated X-ray images of a vehicle obtained by the TPU inspection system with the 9 MeV betatron. The weak vertical bands noticeable on Fig. 6c are due to the X-ray source temporal instability. This inhomogeneity can be eliminated by using a so-called reference channel – a detector(s), in the field of view of which the inspected object does not enter:

$$f_i^{\text{new}} = \frac{f_i^{\text{old}}}{f_i^{\text{ref}}} \times 60000, \quad (9)$$



**Fig. 6** The calibration of the image acquired by detector array: (a) original image, the green outlined part of image was acquired without X-ray; (b) offset calibrated; (c) gain calibrated; (d) reference channel calibrated

where  $f_i^{\text{ref}}$  are the reference channel values and index  $i$  is the image column number. The detectors above the vehicle from Fig. 6c were used as a reference channel to produce calibrated image on Fig. 6d.

## Geometric Transformations

In contrast to the methods considered so far, the geometric transformations change the spatial relationships between the pixels. The geometric transformations are often called transformations of rubber canvas, as they can be imagined as the process of deformation of an image on this canvas.

Figure 7 shows that the bilinear interpolation for the image scaling is preferable than the nearest neighbor interpolation.

Many X-ray systems produce some image geometry distortions (Bukhari and Dailey 2013), which can be corrected by a method of corresponding points. To apply this method, an X-ray image of a special plate with a square grid composed of small holes needs to be acquired. Figure 8b shows this image rectified from unimportant details. Figure 8a shows the reference (undisturbed) image of those holes.

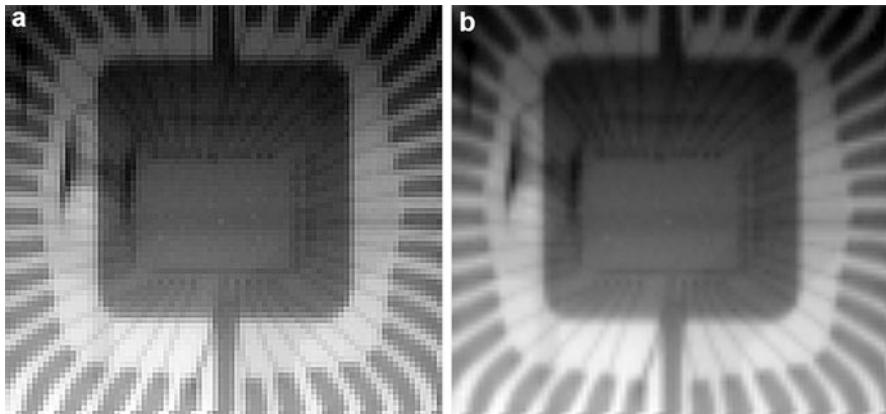
This method splits the current image into a set of triangular elements of type  $B$ , each of which corresponds to a rectangular isosceles triangle  $A$  of the reference image. Next, for each pixel in the triangle  $A$ , the method locates the corresponding point inside the triangle  $B$  by the affine transformation. The gray value at this point is calculated by the four nearest pixels of the current image by the bilinear interpolation.

The formula of the affine transformation of an arbitrary point  $(X, Y)$  of the triangle  $A$  to the point  $(x, y)$  of the triangle  $B$  has the form:

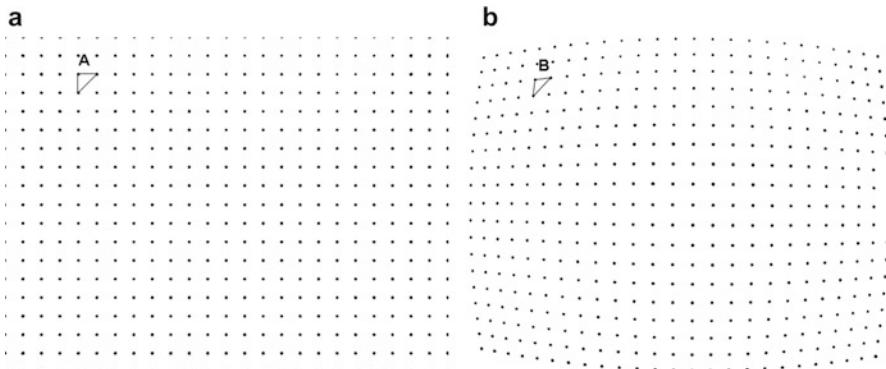
$$\begin{aligned} x &= C_1X + C_2Y + C_3 \\ y &= C_4X + C_5Y + C_6 \end{aligned}, \quad (10)$$

where  $C_i$  ( $i = 1, \dots, 6$ ) are unknown coefficients. The coordinates of the vertices of the triangle  $A$  denote  $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$ ; and the coordinates of the vertices of the triangle  $B$  denote  $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ . Substitution into (10) instead  $(X, Y)$ , the coordinates of the vertices of the triangle  $A$  and instead of  $(x, y)$  the coordinates of the corresponding vertices of the triangle  $B$  allows a system of 6 linear equations to determine  $C_i$ . This system is not ill conditioned, because the vertices of a right-angled isosceles triangle  $A$  do not lie on the same line. Moreover, this system is divided into two independent subsystems for the coefficients  $C_i$  ( $i = 1, 2, 3$ ) and  $C_i$  ( $i = 4, 5, 6$ ), each of which in turn reduces to a system with a triangular matrix by subtracting one of the equations of the other two. The last simplification is due to the fact that the vertices of triangle  $A$  are located at the nodes of a rectangular grid.

Figure 9 shows the image corrected by the corresponding points method and the acquired one.



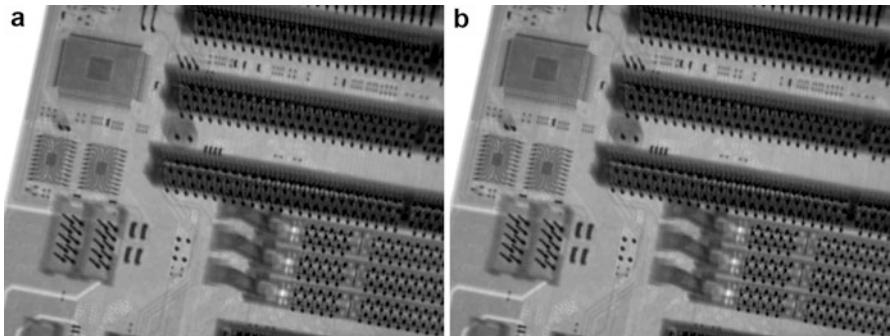
**Fig. 7** Image interpolation: (a) nearest neighbor, (b) bilinear



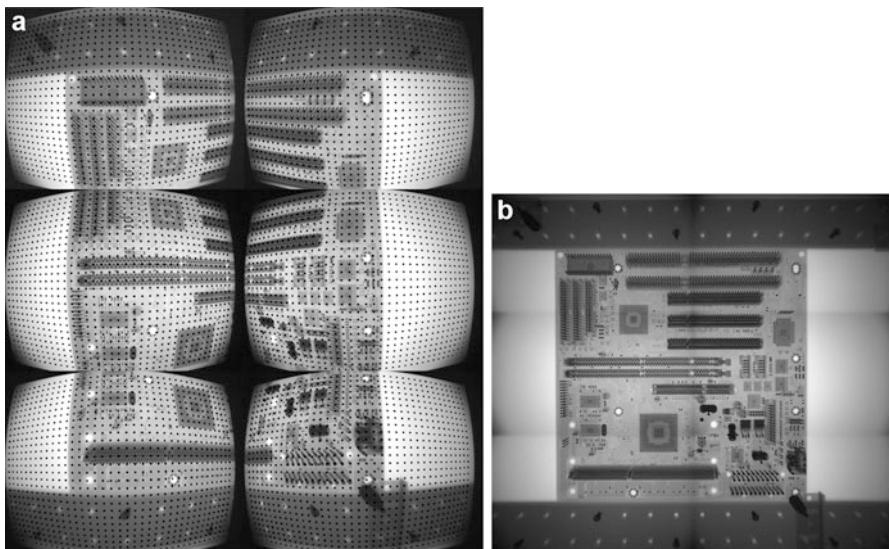
**Fig. 8** Corresponding grid points for the reference and current images. *A* is the reference triangle image element, *B* is the current element, corresponding to *A*

## Image Stitching

The above mentioned method of corresponding points was used for combining multiple X-ray images with overlapping fields of view to produce a segmented panorama (Chakhlov et al. 2006). Figure 10a shows original six X-ray images of mother board with superimposed corresponding points. The stitching of those images (Fig. 10b) is excellent geometrically, but to correct the brightness of panorama, one should apply the background removal (similar that was used on Fig. 4) of the original images.



**Fig. 9** Image correction: (a) corrected image, (b) the original one



**Fig. 10** Image stitching by corresponding points method: (a) 6 original X-ray images of mother board; (b) resulting panorama image without background correction

## Image Filtering

Image filtering (Mery 2015) uses values of pixels from some neighborhood of the input image pixel  $X_{i,j}$  to produce a new gray value  $Y_{i,j}$  in the output image:

$$Y_{i,j} = f(X_{i-m, j-n}, \dots, X_{i,j}, \dots, X_{i+m, j+n}). \quad (11)$$

for  $i = m+1, \dots, M-m$  and  $j = n+1, \dots, N-n$ , where  $M$  and  $N$  are the number of rows and columns of the input and output images. The size of the filter mask in this

case is  $(2m + 1) \times (2n + 1)$ . The operator  $f$  can be linear or nonlinear. In this section, the most important linear and nonlinear filters for X-ray image processing are outlined.

The operator  $f$  is linear, if the resulting value  $Y_{i,j}$  is calculated as a linear combination of the input pixels:

$$Y_{i,j} = \sum_{p=-m}^m \sum_{q=-n}^n h_{p,q} X_{i-p,j-q}, \quad (12)$$

where  $\mathbf{h}$  is called the *convolution mask*. *Averaging* is a simple example of linear filtering. For a  $3 \times 3$  neighborhood, the convolution mask is:

$$\mathbf{h} = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \quad (13)$$

The *median filter* is a ranking operator (and thus nonlinear) where the output value is the middle value of the input values ordered in a rising sequence (Gonzalez and Woods 2008). Figure 11 shows that a  $3 \times 3$  median filter removed the random noise totally while the averaging smeared the noise in the image.

Figure 12 shows the effect of the averaging and median filters with a five-element window on stepped, saw tooth, pulsed, and triangular discrete signals. It can be seen that the median filter does not affect the stepped and saw tooth functions, which is usually a desirable property. However, this filter suppresses impulse signals with length less than half the width of the window. The filter also causes a flattening of the vertex of the triangular function.

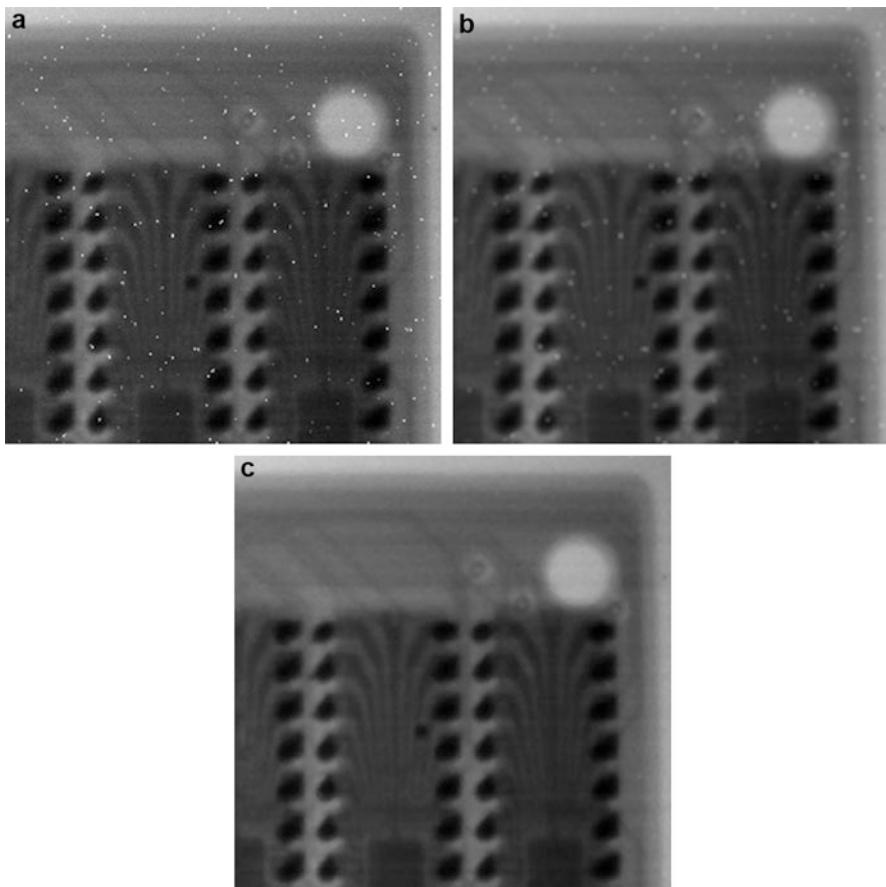
Psychophysical experiments show that an image with underlined boundaries is often subjectively more pleasant than a photometric perfect reproduction. The unsharp mask creates an image that is less blurry than the original. The *unsharp mask* formula is (Gonzalez and Woods 2008):

$$f_{i,j}^{\text{new}} = f_{i,j}^{\text{mean}} + C \times (f_{i,j} - f_{i,j}^{\text{mean}}). \quad (14)$$

The new value is equal to the mean value plus the difference between the current value and the mean multiplied by the constant  $C$ . The  $C$  values are taken usually in the range from 1.5 to 5. Figure 13 shows an X-ray image of a step sample, whose lower part was processed by the unsharp masking  $5 \times 5$  with  $C = 5$ . The processed part, although clearer, may be a less accurate representation of the image's subject.

Combined domain and range filtering is denoted as *bilateral filtering* (Paris et al. 2009; Tomasi and Manduchi 1998). It replaces the pixel value  $X_{i,j}$  with an average of similar and nearby pixel values. The simplest formula for the bilateral filter is:

$$Y_{i,j} = \frac{1}{S} \sum_{p=-m}^m \sum_{q=-n}^n X_{i+p,j+q}, \text{ where } |X_{i+p,j+q} - X_{i,j}| < T. \quad (15)$$



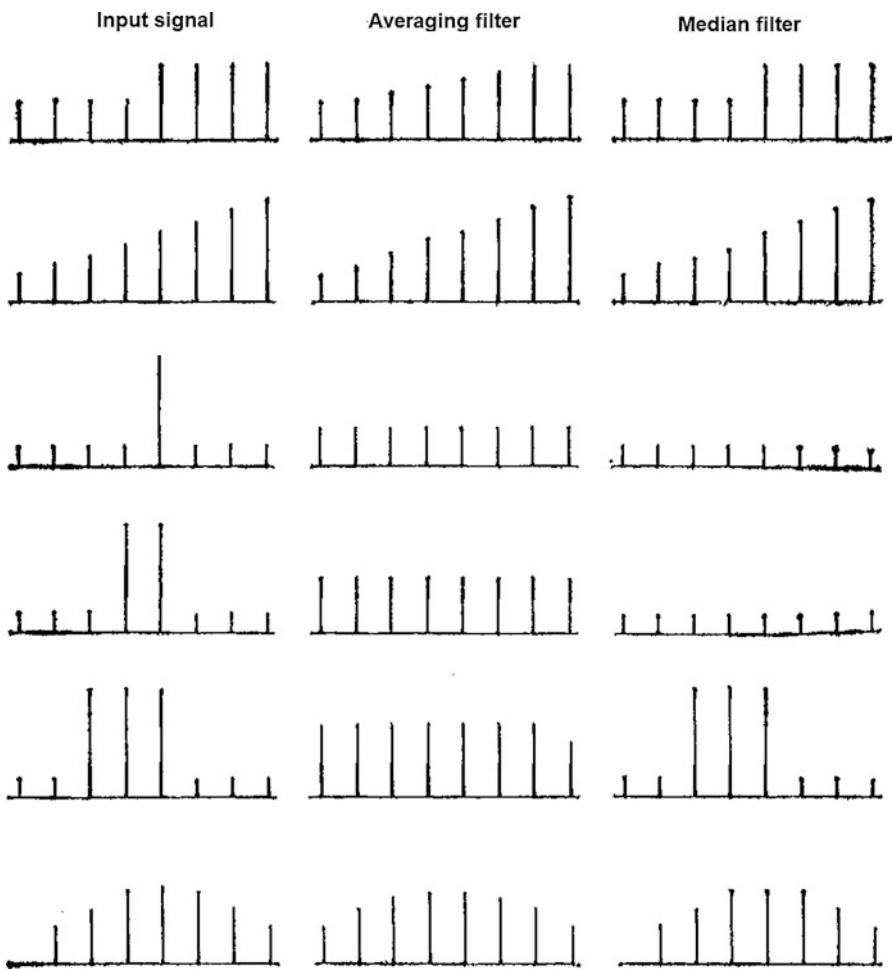
**Fig. 11** Averaging and median filter comparison:  $3 \times 3$ : (a) original image, (b) averaging, (c) median

where  $T$  is a threshold for pixel values,  $S$  is the number of pixels which satisfy the inequality (15). This filter smooths an image while preserving its sharp edges. It is very helpful when preprocessing the dual energy images (Manduca et al. 2009).

---

## Edge Detection

The edges correspond to pixels of the image in which the gray value changes significantly over a short distance (Castleman 1996). Since edges are discontinuities in the intensity of the X-ray image, they are normally estimated by maximizing the gradient of the image. Edge detection image corresponds to a binary image (of the same size of the X-ray image), where a pixel is “0” if it belongs to an edge, otherwise it is “1.”



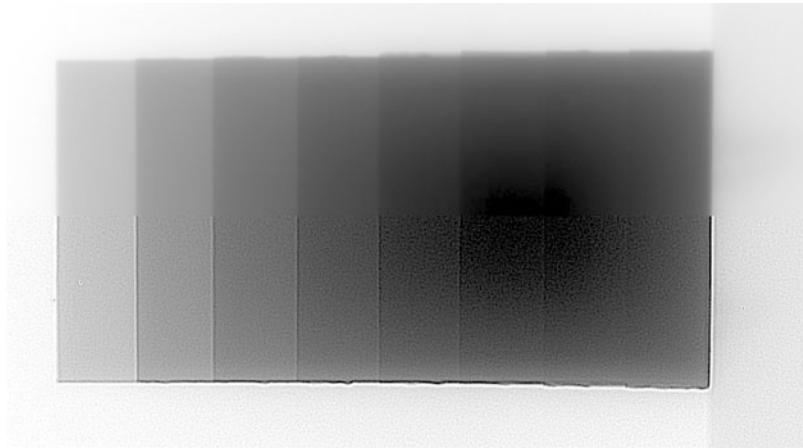
**Fig. 12** 1D-examples of averaging and median filtering  $5 \times 1$

A simple way to calculate the gradient of image  $f_{i,j}$  in the  $i$  and  $j$  directions can be written respectively:

$$\begin{aligned}\frac{\partial f}{\partial x} &\approx f_{i+1,j} - f_{i,j} \\ \frac{\partial f}{\partial y} &\approx f_{i,j+1} - f_{i,j}\end{aligned}. \quad (16)$$

But usually in practice, more complicated filters are used (Gonzalez and Woods 2008). One is due to *Roberts*:

$$\nabla f = |z_9 - z_5| + |z_8 - z_6| \quad (17)$$



**Fig. 13** Unsharp masking applied to the lower part of the image

and another due to *Sobel*:

$$\nabla f = |(z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)| + |(z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)|, \quad (18)$$

where the notation  $\begin{pmatrix} z_1 & z_2 & z_3 \\ z_4 & z_5 & z_6 \\ z_7 & z_8 & z_9 \end{pmatrix}$  is used. Figure 14 shows the results of these filters.

*Canny* proposes a linear mask for edge detection based on an optimization approach (Canny 1986). The idea is to use a derivative of a Gaussian mask to find the local maxima of the gradient of the image. The practical implementation uses adaptive thresholding of the gradient (to detect strong and weak edges) with hysteresis (weak edges are detected only if they are connected to strong edges).

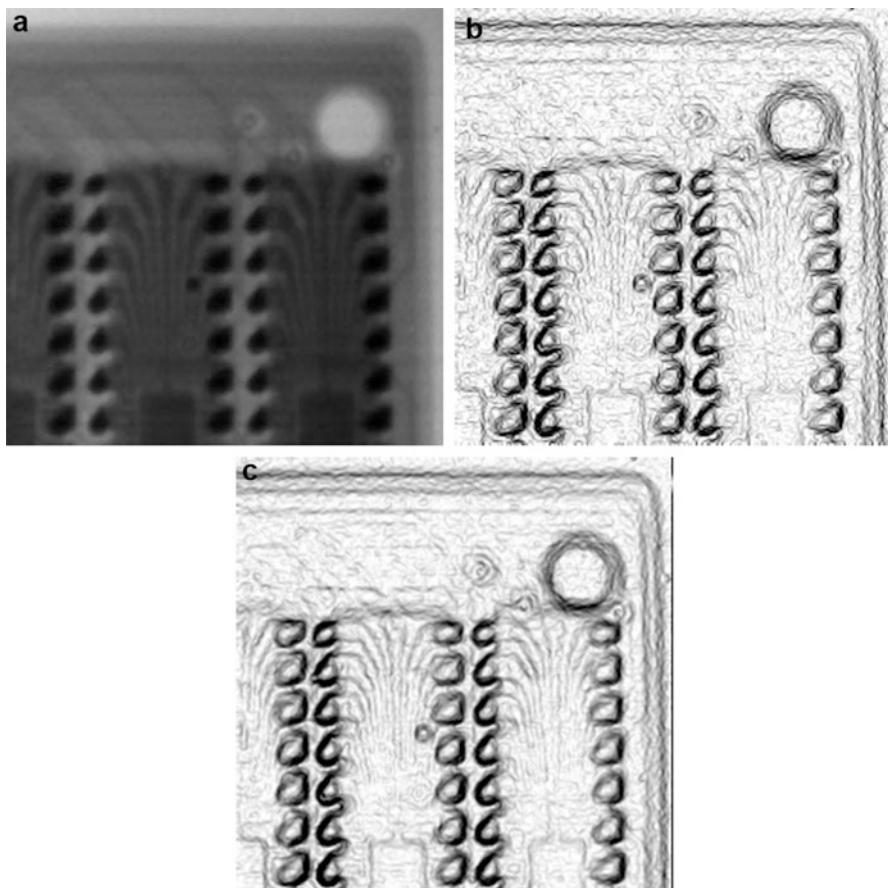
The *emboss* filter is very close to the gradient filters (Gonzalez and Woods 2008):

$$\begin{aligned} f_{i,j}^{\text{new,horiz}} &= A \times (f_{i+1,j} - f_{i-1,j}) + B \\ f_{i,j}^{\text{new,vert}} &= A \times (f_{i,j+1} - f_{i,j-1}) + B, \end{aligned} \quad (19)$$

where  $A$  is an amplitude scale and  $B$  is equal to 32,768 for 16-bpp images. Figure 15 shows the results of these filters for the image from Fig. 14a.

## Segmentation

Segmentation is one of the most difficult processes in image processing. Image segmentation is defined as subdividing an image into disjointed regions (Castleman 1996). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image



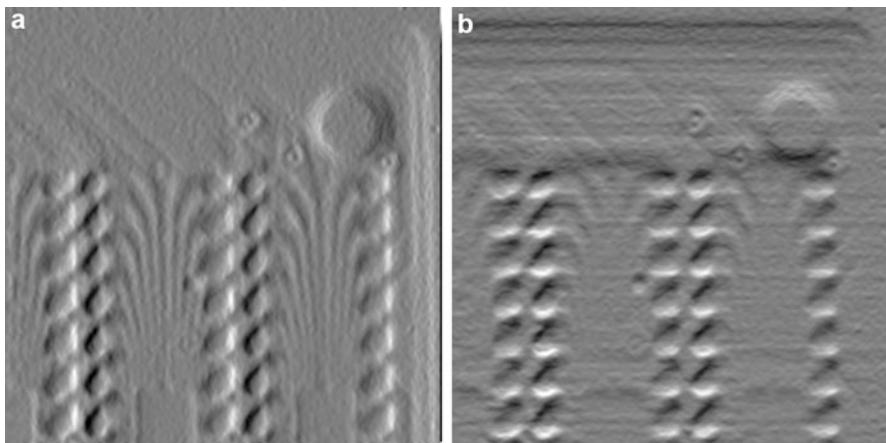
**Fig. 14** Gradient filters: (a) original image, (b) Roberts filter, (c) Sobel filter

segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image so that the pixels with the same label share certain characteristics.

The simplest method of image segmentation is called the *thresholding* method. This method is based on a threshold value to turn a gray-scale image into a binary image. There is also a balanced histogram thresholding. The key of this method is to select the threshold value (or values when multiple-levels are selected). Several popular methods are used in industry including the maximum entropy method, Otsu's method (maximum variance), and k-means clustering (Batenburg and Sijbers 2009).

The *clustering* technique (K-means algorithm) divides an image into K clusters (Barghout and Sheynin 2013). The basic algorithm is:

1. Pick K cluster centers, either randomly or based on some heuristic method, for example K-means++.



**Fig. 15** Emboss filters: (a) horizontal, (b) vertical

2. Assign each pixel in the image to the cluster that minimizes the distance between the pixel and the cluster center.
3. Recompute the cluster centers by averaging all of the pixels in the cluster.
4. Repeat steps 2 and 3 until convergence is attained (i.e., no pixels change clusters).

In this case, distance is the squared or absolute difference between a pixel and a cluster center. The difference is based on pixel gray value and location, or a weighted combination of these factors. K can be selected manually, randomly, or heuristically. This algorithm is guaranteed to converge, but it may not return an optimal solution. The quality of the solution depends on the initial set of clusters and the value of K.

*Edge detection* techniques are used as the base of a segmentation technique. The edges identified by the edge detection are often disconnected. To segment an object from an image however, one needs closed region boundaries (Lindeberg and Li 1997).

*Region-growing* methods are based on the assumption that the neighboring pixels within one region have similar values. The common procedure is to compare one pixel with its neighbors. If a similarity criterion is satisfied, the pixel can be set to belong to the cluster as one or more of its neighbors. The selection of the similarity criterion is significant and the results are influenced by the noise in all instances (Nock and Nielsen 2004).

The central idea of *curve propagation* is to evolve an initial curve towards the lowest potential of a cost function, where its definition reflects the task to be addressed (Caselles et al. 1997). As for most inverse problems, the minimization of the cost functional is nontrivial and imposes certain smoothness constraints on the solution, which in the present case can be expressed as geometrical constraints on the evolving curve.

In *graph partitioning* methods, the image is modeled as a weighted, undirected graph. Usually, a pixel or a group of pixels are associated with nodes and edge

weights defining the similarity between the neighborhood pixels. The image is then partitioned according to a criterion designed to model “good” clusters. Each partition of the pixels output from these algorithms is considered an object segment in the image. Some popular algorithms of this category are normalized cuts (Shi and Malik 2000), random walker (Grady 2006), minimum cut (Wu and Leahy 1993), isoperimetric partitioning (Grady and Schwartz 2006), and minimum spanning tree-based segmentation (Zahn 1971).

Figure 16 shows an example of segmentation process for the welding defects (Carrasco and Mery 2004). First a median filter is used for noise reduction; second, a bottom-hat filter is used to separate hypothetical flaws from their background; third, the segmented regions are identified by means of binary thresholding; fourth, filters taken from morphological mathematics are used to eliminate over-segmentation; and fifth the Watershed transform is used to separate internal regions.

---

## Computer Vision Problems (Features Extraction and Selection for Classification)

The segmentation is used to detect regions that can be the objects of interest. As segmented potential regions frequently set off false detections, an analysis of the segmented regions can significantly improve the effectiveness of detection. Measuring certain characteristics of the segmented regions (feature extraction) can help to distinguish the false detection. Therefore, a feature selection must be performed. Depending on the values returned for the selected features, one can try to classify each segmented potential region in one of the following two classes: background or object of interest.

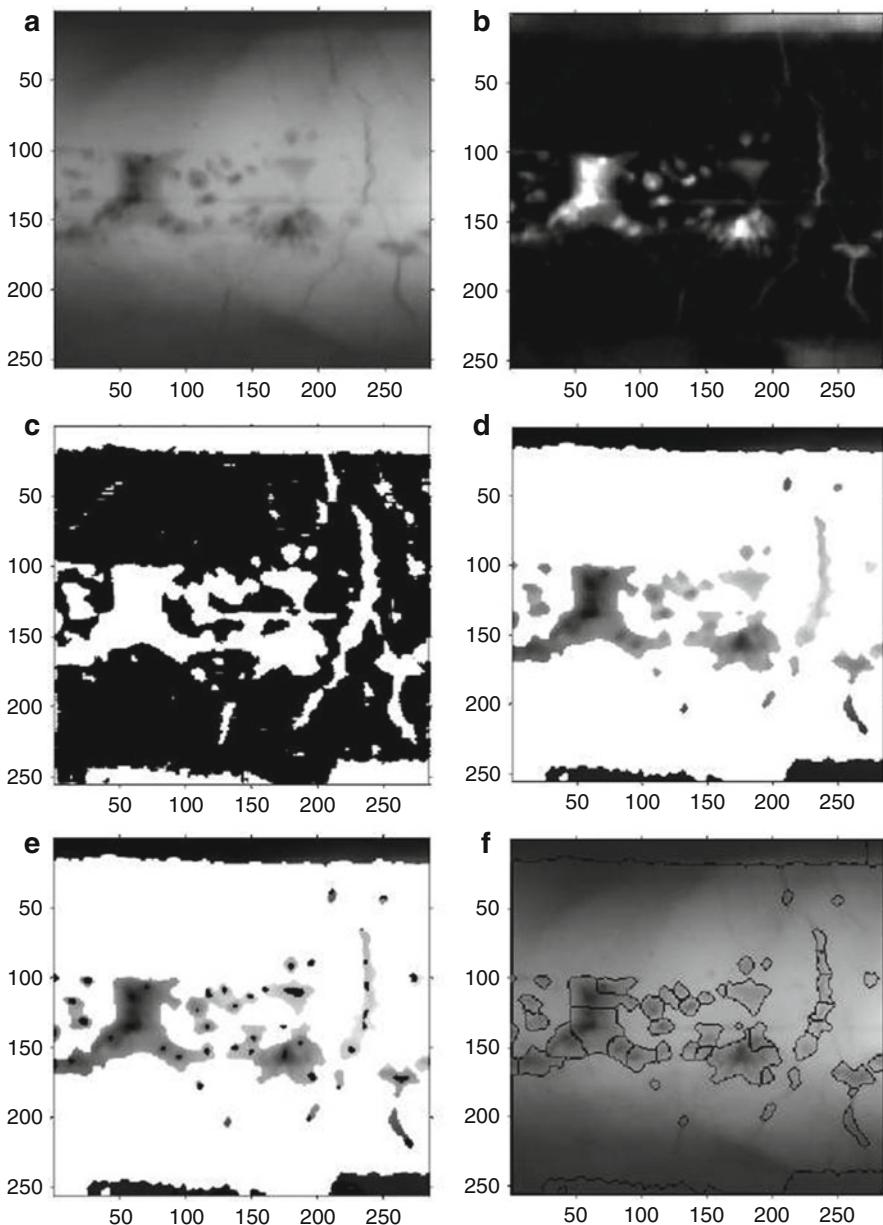
The features are divided into two groups: geometric (height and width, area and perimeter, center of mass, roundness, et al.) and intensity (mean gray value, contrast, intensity moments, crossing line profiles statistical textures, Gabor, filter banks, etc.) features.

The design of a computer vision system (Mery 2015) is shown in Fig. 17: automatically extract and select from a large set of image features and create a bank of classifiers and evaluation of a test image.

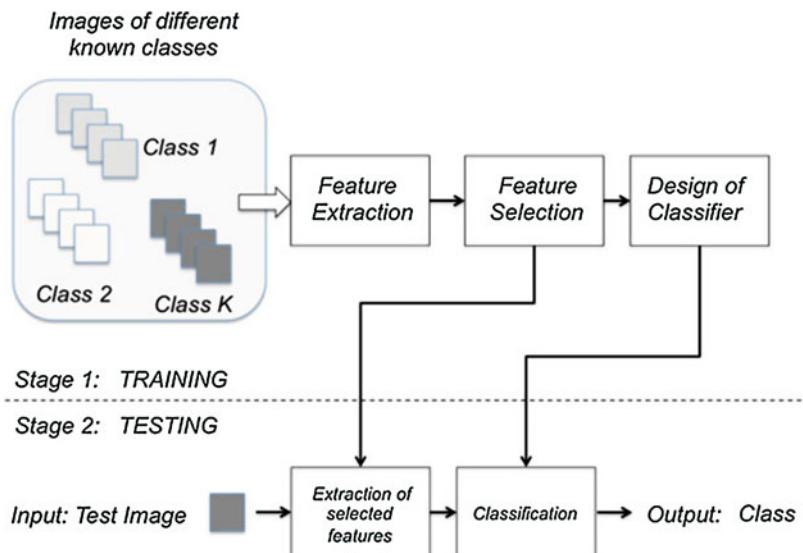
Feature selection is the process of selecting a subset of relevant features for use in model construction. Feature selection techniques are used for four reasons:

- Simplification of models to make them easier to interpret by researchers
- Shorter training times
- To avoid the curse of dimensionality,
- Enhanced generalization by reducing overfitting (formally, reduction of variance)

The central premise when using a feature selection technique is that the data contains some features that are either redundant or irrelevant and can thus be removed without incurring much loss of information.



**Fig. 16** Example of the segmentation process: (a) image after the median filter, (b) the Bottom-Hat filter, (c) binary thresholding, (d) sectioning process, (e) Modification of minima, (f) the Watershed transformation



**Fig. 17** A computer vision system. In the training stage, features are extracted and selected. In addition, a classifier is designed. In the testing stage, selected features are extracted and the test image is classified

The features can be selected by following methods (Bishop 2006): exhaustive search, branch and bound, sequential backward selection, ranking by class separability criteria, forward orthogonal search, least square estimation, combination with principal components, feature selection based on mutual information.

The most popular classifiers are minimal distance, Mahalanobis distance, Bayes, linear discriminant analysis, quadratic discriminant analysis, k-nearest neighbors, neural networks, support vector machines, and sparse representation.

## Dual Energy Imaging

Conventional X-ray imaging represents the observed object in terms of product of attenuation by material thickness. This information is not sufficient to characterize precisely the observed object. In the energy range less than 300 KeV, the attenuation for X-ray radiation is a combination of two photon-matter interactions: the photoelectric effect and Compton scatter. For energies above 1 MeV, most significant effects are Compton scatter and effect of pair creation.

The two interactions and their relative contribution to the total attenuation are energy dependent. Thus, measurements at two distinct energies should permit the separation of the attenuation into its basic components, which can be used to identify material, and finally to produce material specific image. Let remind the principles of dual-energy technique (Rebuffel and Dinten 2007).

If a monochromatic source emits a number of photons  $N_0$ , then the number of photons  $N$  imaging on a collimated detector of one pixel (linear response) after attenuation by a thickness  $T$  (cm) of an object of attenuation  $\mu$  ( $\text{cm}^{-1}$ ) is given by:  $N = N_0 \exp(-\mu \cdot T)$ .

The coefficient  $\mu$  depends on the density of the material ( $\rho$ ), on its chemical composition (effective atomic number  $Z$ ), and on the energy  $E$  of the photons:  $\mu = \rho \cdot \tau(E, Z)$  where  $\tau$  is the mass attenuation of the material ( $\text{cm}^2/\text{g}$ ).

If we compute the log-measurement of the attenuation, we get:  $m_E = -\log(N/N_0) = T \cdot \mu = T \cdot \rho \cdot \tau(E, Z)$ , which is proportional to the thickness crossed by the X-ray. If the object is composed of several materials,  $m_E = \sum_{\text{material } i} T_i \cdot \rho_i \cdot \tau_i(E, Z_i)$ .

Let us now consider two materials 1 and 2. Two acquisitions at two distinct low energy (LE) and high energy (HE) then provide a linear system:

$$\begin{cases} m_{LE} &= \mu_1^{LE} \cdot T_1 + \mu_2^{LE} \cdot T_2 \\ m_{HE} &= \mu_1^{HE} \cdot T_1 + \mu_2^{HE} \cdot T_2 \end{cases} \quad (20)$$

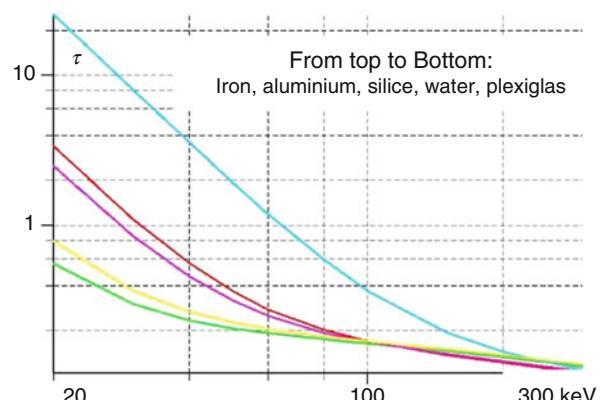
which can be easily solved as long as its determinant  $\Delta = \mu_1^{LE} \cdot \mu_2^{HE} - \mu_1^{HE} \cdot \mu_2^{LE}$  is non-null, which traduce the fact that chemical properties of the materials differ significantly.

Figure 18 gives examples of values of  $\tau(E)$  for some materials.

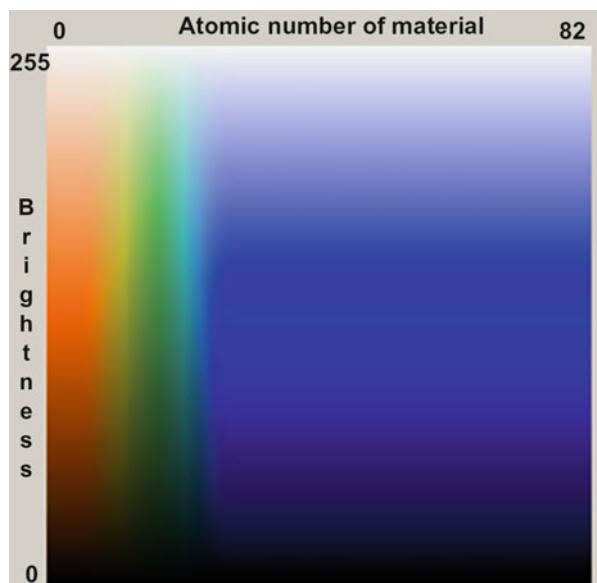
Most of the methods rely on the hypothesis that the attenuation coefficient can be decomposed linearly on a basis of two functions  $\alpha(E)$  and  $\beta(E)$ , depending only on energy of emitted photons, and corresponding to photoelectric and Compton effects – for instance:  $\mu(\rho, Z, E) = \rho Z^n \alpha(E) + \rho \beta(E)$ , where the exponent  $n$  varies between 3 and 4.

To paint images of dual energy method, there are used two matrices: the atomic number distribution and one of the original matrices of brightness for LE or HE. If there is used matrix of HE brightness, than talking about dual image with high penetration, and vice versa dual image with low penetration. Also there is used a special matrix of colors (Fig. 19). For organic materials with atomic number less 12 – colors are orange-brown. For metal with atomic number more than 23 – colors

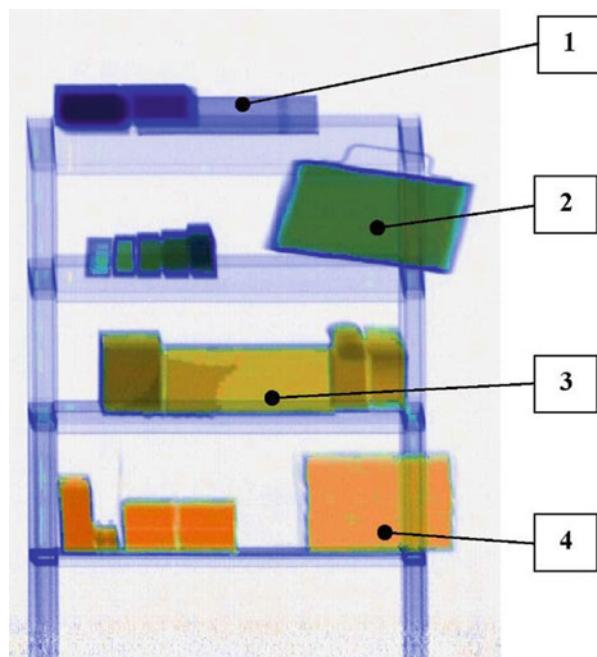
**Fig. 18** Mass attenuation functions for some materials, from 20 to 300 keV



**Fig. 19** Colors to paint image of dual energy method



**Fig. 20** Dual-energy imaging for betatron X-ray source ( $LE = 4.5$  MeV,  $HE = 7.5$  MeV): 1 lead, 2 iron, 3 aluminum, 4 plexiglas



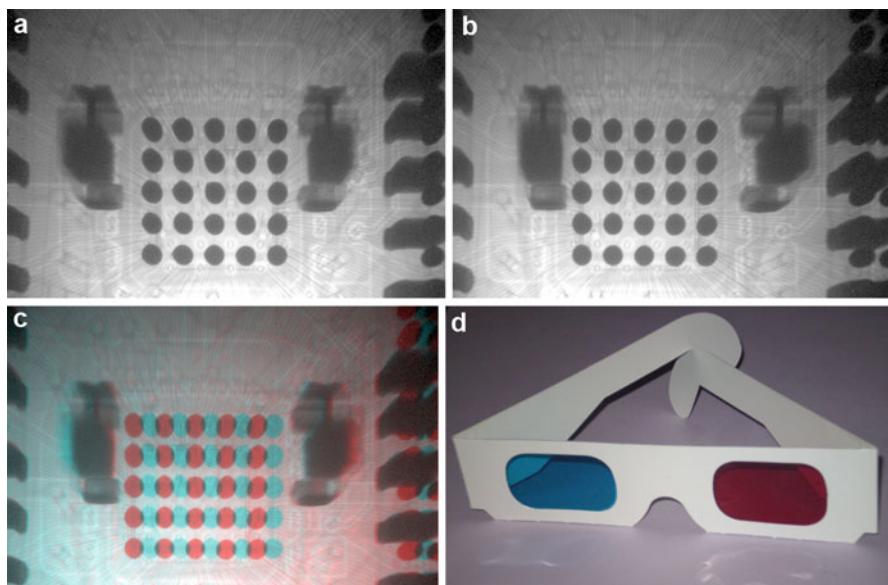
are blue. And for range from 12 to 23, corresponds to material mix – colors are green. Thus to paint arbitrary image point, one can select color, which correspond to the intersection of column  $i$  (atomic number of selected point) and line  $j$  (brightness of selected point for  $LE$  or  $HE$ ).

Figure 20 shows the dual energy imaging of test objects from lead, iron, aluminum, and plexiglas (Chakhlov et al. 2016) for betatron X-ray source:  $LE = 4.5$  MeV,  $HE = 7.5$  MeV.

### Stereoscopic 3D Images

The use of X-ray images of a test object from two different directions makes it possible to take stereoscopic 3D X-ray images. In contrast to conventional planar X-ray images, with stereoscopic 3D images the operator sees the test object as a transparent, three-dimensional object with spatial depth. The stereoscopic images can be presented on special 3D monitors or by anaglyph images.

Anaglyph 3D images contain two differently filtered colored images, one for each eye. When viewed through the cyan and red “anaglyph glasses,” each of the two images reaches the eye it’s intended for, revealing an integrated stereoscopic image. Figure 21 shows two original X-ray images, the resulting anaglyph image and cyan-red paper glasses.



**Fig. 21** Creation anaglyph X-ray image: (a) left X-ray image, (b) right X-ray image, (c) resulting anaglyph image (d) cyan and red anaglyph glasses

## Summary

In this chapter, we covered the main techniques of X-ray image processing. They are: contrast enhancement; background removal; frame averaging; offset, gain, and reference channel calibrations; geometric transformations; image stitching, image filtering; edge detection; segmentation, features extraction and selection for classification, dual energy imaging, anaglyph 3D images.

The chapter provided a short overview, presenting the most popular methodologies with examples and the use of actual X-ray images.

---

## References

- ASTM standard E 2339–08 (2008) Standard practice for digital imaging and communication in nondestructive evaluation (DICONDE). ASTM international, West Conshohocken, PA, <http://www.astm.org>, Accessed 3 Oct 2018
- Barghout L, Sheynin J (2013) Real-world scene perception and perceptual organization: lessons from computer vision. *J Vision* 13(9):709–709. <https://doi.org/10.1167/13.9.709>
- Batenburg J, Sijbers J (2009) Adaptive thresholding of tomograms by projection distance minimization. *Pattern Recogn* 42(10):2297–2305. <https://doi.org/10.1016/j.patcog.2008.11.027>
- Bishop CM (2006) Pattern recognition and machine learning. Springer, New York 738p
- Bukhari F, Dailey N (2013) Automatic radial distortion estimation from a single image. *J Math Imag Vis* 45(1):31–45. <https://doi.org/10.1007/s10851-012-0342-2>
- Canny J (1986) A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell PAMI* 8(6):679–698. <https://doi.org/10.1109/TPAMI.1986.4767851>
- Carrasco M, Mery D (2004) Segmentation of welding defects using a robust algorithm. *Mater Eval* 62(11):1142–1147. [https://doi.org/10.1142/9789812770943\\_0027](https://doi.org/10.1142/9789812770943_0027)
- Caselles V, Kimmel R, Sapiro G (1997) Geodesic active contours. *Int J Comp Vision* 22(1):61–79. <https://doi.org/10.1023/A:1007979827043>
- Castleman K (1996) Digital image processing. Prentice-Hall, Englewood Cliffs
- Chakhlov S, Lebedev M, Usachev E (2006) Method of X-ray image stitching. *Testing Diagnostics*, 2:34–40 (Russian)
- Chakhlov S, Kasyanov S, Kasyanov V, Osipov S, Stein M, Stein A, Sun X (2016) Betatron application in mobile and relocatable inspection systems for freight transport control. *J Phys Conf Ser* 671:012024 IOP Publishing. <https://doi.org/10.1088/1742-6596/671/1/012024>
- Gonzalez R, Woods R (2008) Digital image processing, 3rd edn. Prentice Hall, Upper Saddle River
- Grady L (2006) Random walks for image segmentation. *IEEE Trans Pattern Anal Mach Intel* 28(11):1768–1783. <https://doi.org/10.1109/TPAMI.2006.233>
- Grady L, Schwartz E (2006) Isoperimetric graph partitioning for image segmentation. *IEEE Trans Pattern Anal Mach Intel* 28(3):469–475. <https://doi.org/10.1109/TPAMI.2006.57>
- Hanes R (1949) The construction of subjective brightness scales from fractionation data: a validation. *J Experiment Psych* 39(5):719–728. <https://doi.org/10.1037/h0053962>
- ISee! (2017) BAM radiographic image analysis software. <http://vision-in-x.com/en/> Accessed 3 Oct 2018
- Jobst M, Koetz A, Clendening S (2010) The value of diconde in multi-modal NDT environments. In: ECNDT 2010, pp 3912–3917
- Lindeberg T, Li X (1997) Segmentation and classification of edges using minimum description length approximation and complementary junction cues. *Comp Vision Image Understand* 67(1):88. <https://doi.org/10.1006/cviu.1996.0510>

- Manduca A, Yu L, Trzasko J, Khaylova N, Kofler J, McCollough C, Fletcher J (2009) Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT. *Med Phys* 36(11):4911–4919. <https://doi.org/10.1118/1.3232004>
- Mery D (2015) Computer vision for X-ray testing (imaging, systems, image databases and algorithms). Springer, Berlin 362p. <https://doi.org/10.1007/978-3-319-20,747>
- Mery D (2018) BALU: a toolbox Matlab for computer vision, pattern recognition and image processing. <http://dmery.ing.puc.cl/index.php/balu>. Accessed 3 Oct 2018
- Mery D, Riffó V, Zschepel U, Mondragón G, Lillo I, Zuccar I, Lobel H, Carrasco M (2015) GDXRay: the database of X-ray images for nondestructive testing. *J Nondestruct Eval* 34(4):42. <https://doi.org/10.1007/s10921-015-0315-7>
- Nock R, Nielsen F (2004) Statistical region merging. *IEEE Trans Pattern Anal Mach Intel* 26(11):1452–1458. <https://doi.org/10.1109/TPAMI.2004.110>
- Paris S, Kornprobst P, Tumblin J, Durand F (2009) Bilateral filtering: theory and applications. *Found Trends Comput Graph Vis* 4(1):1–73. <https://doi.org/10.1561/0600000020>
- Pratt W (2007) Digital image processing, 4th edn. Wiley, New York
- Rakshit S, Ghosh A, Uma Shankar B (2007) Fast mean filtering technique (FMFT). *Pattern Recogn* 40:890–897. <https://doi.org/10.1016/j.patcog.2006.02.008>
- Rebuffel V, Dinten J-M (2007) Dual-energy X-ray imaging: benefits and limits. *Insight Nondestruct Test Cond Monitor* 49(10):589–594. <https://doi.org/10.1784/insi.2007.49.10.589>
- Shi J, Malik J (2000) Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Machine Intel* 22(8):888–905. <https://doi.org/10.1109/34.868688>
- Tomasi C, Manduchi R (1998) Bilateral filtering for gray and color images. In: Proceedings of the 1998 I.E. international conference on computer vision, Bombay, India, pp. 839–846, <https://doi.org/10.1109/ICCV.1998.710815>
- Wu Z, Leahy R (1993) An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. *IEEE Trans Pattern Anal Mach Intel* 15(11):1101–1113
- Zahn C (1971) Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Trans Comp* 20(1):68–86
- Zuiderveld K (1994) Contrast limited adaptive histogram equalization. In: Heckbert P (ed) *Graphics gems IV*. Academic Press, New York, pp 474–485. <https://doi.org/10.1016/B978-0-12-336156-1.50061-6>



# X-Ray Phase Contrast Methods

30

Sheridan Mayo and Marco Endrizzi

## Contents

Introduction .....	1054
A Brief History of Conventional Absorption-Based X-Ray Imaging .....	1054
From Absorption-Contrast to Phase-Contrast Imaging .....	1055
Phase-Contrast Imaging .....	1058
Interferometry .....	1058
Analyzer-Based Imaging (ABI) .....	1058
Propagation-Based Phase-Contrast Imaging (PB PCI) .....	1061
Grating-Based Phase-Contrast Imaging (GB PCI) .....	1071
Emerging Techniques .....	1077
Summary .....	1083
References .....	1083

## Abstract

X-ray phase-contrast methods for imaging and tomography have considerable advantages over conventional absorption contrast. They enable excellent imaging contrast for both high- and low-density materials within the same sample and offer additional imaging modes which can highlight low-contrast boundaries, subtle density gradients, and fine-scale texture. Phase-contrast methods encompass a family of techniques implemented both in the laboratory and at synchrotron sources which utilize a range of tools to make visible the phase shift imposed

---

S. Mayo (✉)

CSIRO Manufacturing, Clayton, VIC, Australia  
e-mail: [Sherry.Mayo@csiro.au](mailto:Sherry.Mayo@csiro.au)

M. Endrizzi

University College London, London, UK  
e-mail: [m.endrizzi@ucl.ac.uk](mailto:m.endrizzi@ucl.ac.uk)

on an X-ray beam by the sample. This chapter describes the main phase-contrast techniques used, their particular strengths, and a range of applications for each, together with some of the mathematical methods used in the analysis of phase-contrast data.

---

## Introduction

### A Brief History of Conventional Absorption-Based X-Ray Imaging

Within a few years of Roentgen's discovery of X-rays (Roentgen 1896), there was a rapid uptake of X-ray imaging for medical applications. As Roentgen so memorably demonstrated with the image of his wife's hand, the X-ray imaging technique revealed the body's internal structure enabling the diagnosis of broken bones. Although medical applications dominated, as early as 1898, there were experiments using X-ray imaging to examine binary alloys (Heycock and Neville 1898). Within a few decades, the development of higher-energy X-ray tubes opened the door to the use of X-ray imaging for industrial and materials science applications (Coolidge 1917). An early pioneer of the method was H.H. Lester who began his work in radiography of metals in 1920. By the 1930s companies such as Richard Seifert & Co (Germany), GE (USA), and Shimadzu (Japan) were supplying commercial systems capable of radiography of welds and thick steel parts.

As nondestructive imaging (NDI) expanded, microfocus X-ray sources became increasingly important to enable imaging at higher resolutions (Parish 1986). Many significant developments paving the way for this technology were made by Coolidge and his team at GE, including rotating anodes and electron-focussing optics. Micro-focus sources with sizes ranging from several microns to submicron in size are now widely available for NDI applications utilizing both imaging and tomography (Sasov et al. 2001; Stock 2008). More details on X-ray tomography can be found in ► Chap. 32, “3D X-Ray Tomography: Basics and Latest Developments.”

In parallel with laboratory-based techniques and tools, synchrotron radiation sources based at major facilities emerged, offering very powerful X-ray beams. This opened up the ability to use a range of X-ray optics to achieve high-resolution imaging and micro-CT. Some of these techniques employed optical elements and phase-contrast methods analogous to those used in visible light optics for nanoscale imaging (Schmahl et al. 1994). At the macroscopic scale, a series of phase-contrast techniques emerged, better suited to X-ray imaging of larger samples and which were soon translated to laboratory X-ray inspection systems (Endrizzi 2018).

This chapter gives an overview of these phase-contrast imaging techniques, with the major emphasis being on those methods which have been implemented in the lab and are suited to nondestructive imaging (NDI). This includes a review in each case of practical applications of these methods to a range of imaging and tomography problems.

## From Absorption-Contrast to Phase-Contrast Imaging

The common feature of the vast majority of X-ray imaging-based NDI until recent decades is the reliance on absorption contrast in image formation. Just as denser bones are highlighted in a medical X-ray image, images of industrial and materials science samples reveal the X-ray attenuation of the different components. The absorption of X-rays is determined by the linear attenuation coefficient which in turn depends on density and elemental composition. This is typically expressed as Beer's law

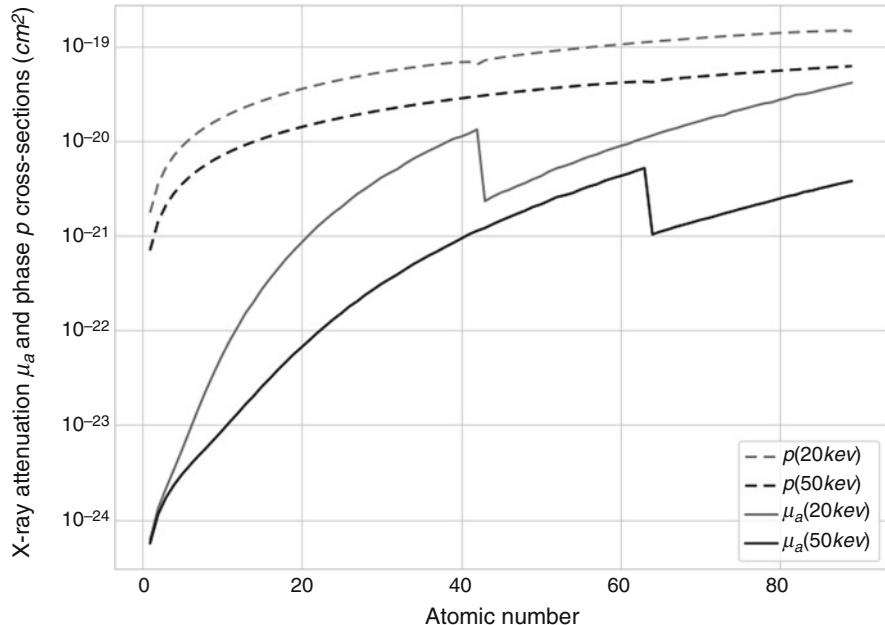
$$I = I_0 e^{-\int \mu(z) dz}$$

where  $I$  is the measured image intensity,  $I_0$  is the illumination without the sample,  $\mu(z)$  is the linear attenuation coefficient in  $\text{cm}^{-1}$  of the sample at point  $z$ , and  $z$  is the position along the propagation direction through the sample. For a homogenous material of thickness  $t$  and with linear attenuation coefficient  $\mu$ , this simplifies to  $I = I_0 e^{-\mu t}$ . Beer's law only applies to monochromatic X-ray beams, where all the X-rays are of a single energy (or wavelength). It outlines the exponential reduction of the beam intensity as the X-ray beam traverses a sample. In laboratory systems the beam is typically polychromatic, made up of a range of X-ray energies; nevertheless the image formation relies on the same principle as Beer's law but summing over a range of X-ray energies and with attenuation coefficients for the sample material that vary with energy.

Successful absorption-contrast imaging depends on the X-ray energies, sample thickness, and attenuation coefficients of the sample material being such that a significant proportion of the X-rays are absorbed to form an image but that sufficient X-rays can pass through the most absorbing part of the sample that the full sample is imaged adequately. The X-ray energy being used can be tuned to optimize imaging by increasing the acceleration voltage of the X-ray tube (in the case of lab sources) and in some cases using filters to remove soft X-rays.

The linear attenuation coefficient of an element increases with atomic number,  $Z$ , and density. For compound materials the linear attenuation coefficient is calculated as a weighted sum of the mass attenuation coefficients of the component elements. Consequently absorption-based X-ray imaging will readily image reasonably dense and higher  $Z$  materials such as metals but will fare less well with low- $Z$  materials. The X-ray energy spectrum can be tailored to a lower-energy range to improve contrast with low- $Z$  materials; however, if there are both high- and low- $Z$  components within the same sample, imaging both adequately will be very difficult using absorption contrast alone.

Attenuation, however, is not the only effect of matter on X-rays. There is also a phase shift or refraction component to the interaction. The plot below shows the relative size of the attenuation and phase cross sections versus  $Z$  at two X-ray energies, 20 keV and 50 keV. The phase cross section drops off much less than attenuation cross section at low  $Z$ , suggesting the possibility of utilizing phase effects for imaging low- $Z$  materials (Fig. 1).



**Fig. 1** Phase and attenuation cross sections for 20 keV and 50 keV X-rays versus atomic number. The step changes seen in the attenuation cross sections occur where that particular X-ray energy is no longer sufficient to eject the inner shell electrons for elements of that atomic number and above

As noted above, X-rays can be both absorbed and refracted on passing through a sample much like visible light. The complex refractive index,  $n$ , is given as

$$n = 1 - \delta - i\beta,$$

where the imaginary part,  $\beta$ , is related to the linear attenuation coefficient,  $\mu$  ( $\text{cm}^{-1}$ ), mentioned in the previous section, such that

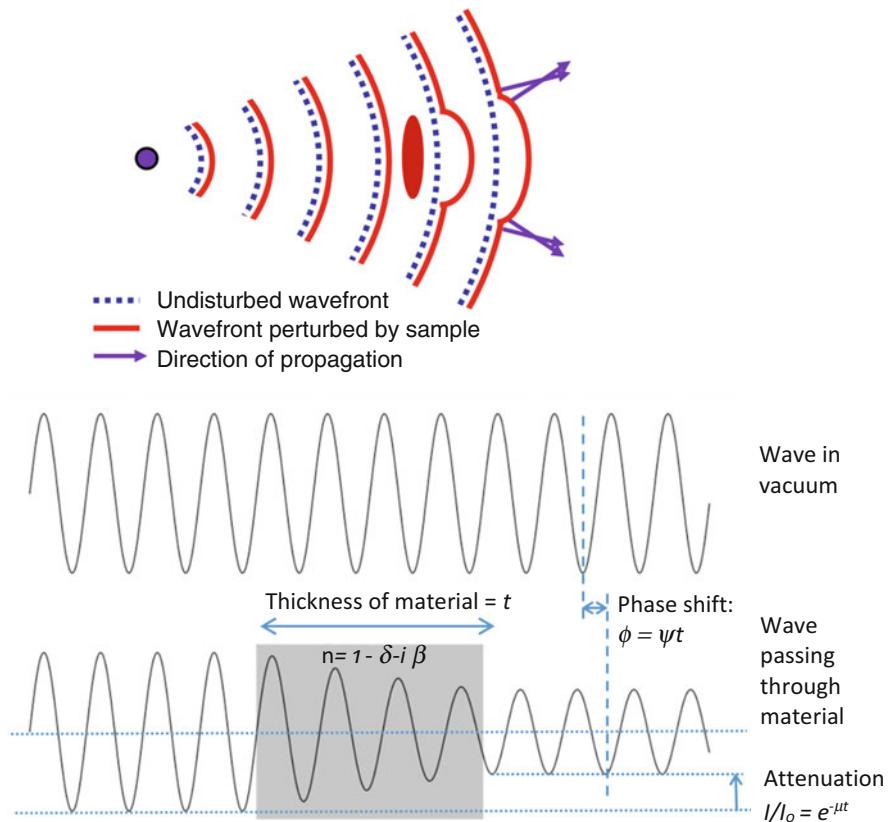
$$\mu = 4\pi\beta/\lambda,$$

where  $\lambda$  is the wavelength of the X-rays (expressed by convention in cm).

In a similar way, the real part,  $\delta$ , of the refractive index corresponds to the linear phase-shift coefficient  $\psi$  ( $\text{cm}^{-1}$ )

$$\psi = -2\pi\delta/\lambda.$$

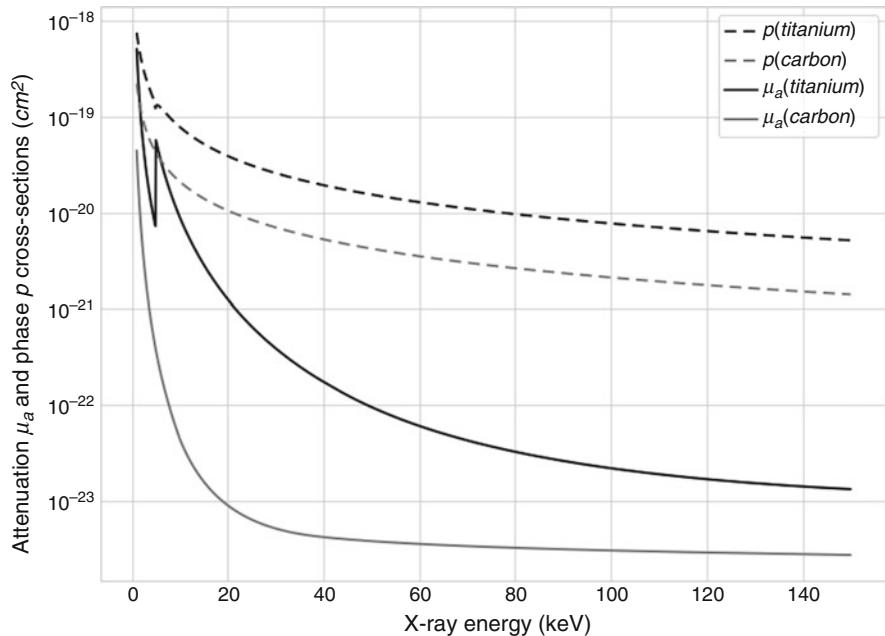
Figure 2 shows the effect of refraction on a spherical wave in the upper image (scale of shift is greatly exaggerated for clarity). The wave front is distorted in proportion to the optical path length through the object. For the part of the wave passing through a thickness  $t$ , this corresponds to a phase shift of  $\phi = \psi t$ . The propagation of the wave is perpendicular to the wave front so that the wave front



**Fig. 2** Effects of a sample on the beam. Upper image shows a spherical wave passing through a sample with resulting wave front distortion and corresponding change in wave propagation direction which can lead to interference (the size of the wave front distortion is greatly exaggerated). Lower image: comparison of a wave in vacuum and a wave passing through a medium of refractive index  $n$ ; the latter shows phase shift and attenuation arising from the real and imaginary parts of the refractive index of the material.

distortion results in corresponding small changes in propagation direction of rays passing through different parts of the sample. The lower part of Fig. 2 compares a wave passing through thickness  $t$  of the material with a reference wave moving in a vacuum and indicates both the attenuation and phase-shift effects arising from the complex refractive index of the material.

Both coefficients of the refractive index are dependent on the density of the material and vary with X-ray energy and  $Z$ , the atomic number. However, not only is the phase shift relatively stronger for low- $Z$  materials than attenuation, it also decreases more slowly with increasing X-ray energy (see Fig. 3). This combination of features opens up the potential to image a wider range of materials at a wider range of energies, including low- and high-density materials simultaneously, provided the effect of the phase shift can be made visible.



**Fig. 3** Phase and attenuation cross sections for carbon and titanium versus X-ray energy

## Phase-Contrast Imaging

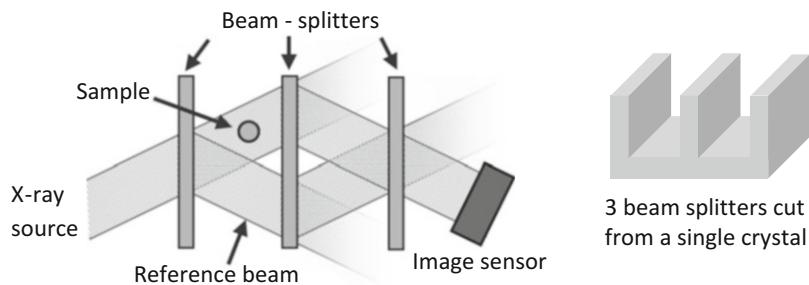
### Interferometry

The earliest attempt to visualize phase contrast was the interferometry approach of Bonse and Hart (1965) who utilized a series of three parallel Laue diffraction beam splitters, produced by cutting two wide slots in a single crystal of silicon as shown in Fig. 4. This setup was used with copper K $\alpha$  X-rays, exploiting the 220 Bragg reflection to split the beam at the first crystal, with the sample placed in one beam path, splitting both beams again at the second crystal with two of the sub-beams able to recombine and interfere at the third crystal. The interference with the “reference beam” enabled the phase shift of the sample beam to be made visible in the form of Moire fringes. This approach was used for imaging low-density materials (Ando and Hosoya 1972).

### Analyzer-Based Imaging (ABI)

#### ABI Fundamentals

Another early approach to phase contrast, known as analyzer-based imaging, makes use of a monochromatic X-ray beam and an “analyzer” crystal monochromator after the sample which acts as an angular filter such that refractive effects caused by the



**Fig. 4** Sketch of Bonse-Hart interferometer layout (left) and 3D view showing how three beam splitters are cut from a single crystal

object are converted into intensity effects in the detector plane. Goetz et al. (1979) first developed this method and applied it to imaging the internal structure of pellets for use in thermonuclear fusion experiments. These and similar early experiments used laboratory X-ray sources and clearly demonstrated enhanced contrast in a variety of specimens; however, it has been more widely employed at synchrotron sources with the major focus being on biological or biomedical applications.

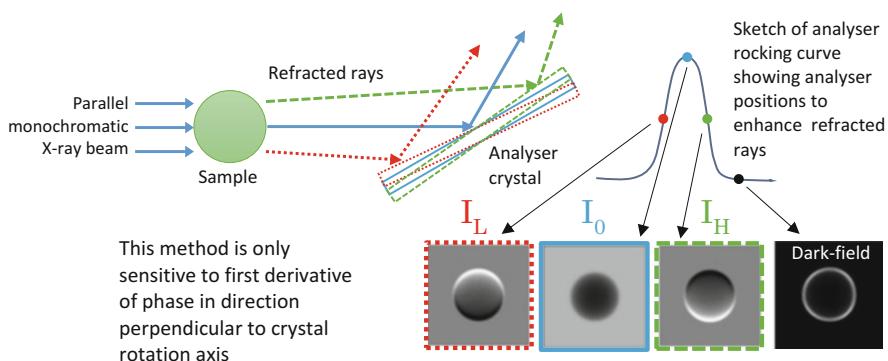
On passing through a sample, X-rays are refracted at an angle relative to the unrefracted beam which is proportional to the local gradient of the phase shift imposed by the corresponding part of the sample (Born and Wolf 1980)

$$\Delta\theta = \frac{\lambda}{2\pi} \frac{\partial\phi}{\partial x}$$

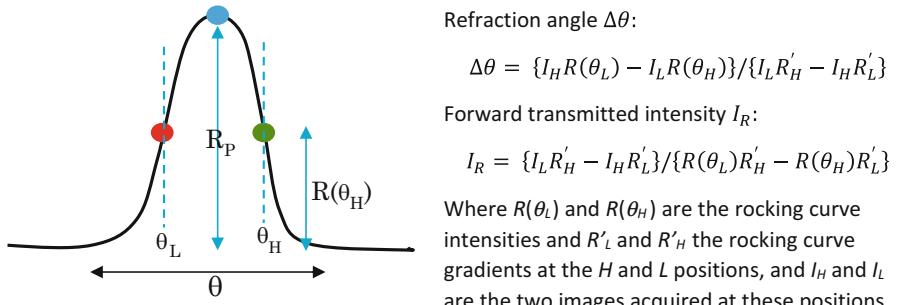
The analyzer crystal will only diffract the unrefracted X-ray beam when orientated to it at the Bragg angle, which depends on the crystal planes being used and the X-ray energy. However, by rocking the crystal slightly off the Bragg angle, the refracted beams (for which the new position better satisfies the Bragg condition) are more efficiently reflected by the crystal. A sketch showing the key elements of the setup is shown in Fig. 5.

The images acquired with the crystal positioned on the sides of the rocking curve (corresponding to  $I_H$  and  $I_L$  in Fig. 5) contain absorption-contrast information but are also sensitive to the first derivative of the phase shift in the direction perpendicular to the crystal rotation axis (see Fig. 11). The refraction and absorption components can be extracted from these images as shown by Chapman et al. (1997) with their “DEI” or diffraction-enhanced imaging approach (see Fig. 6). The actual phase shift produced by a sample is calculated by integrating the  $\Delta\theta$  image.

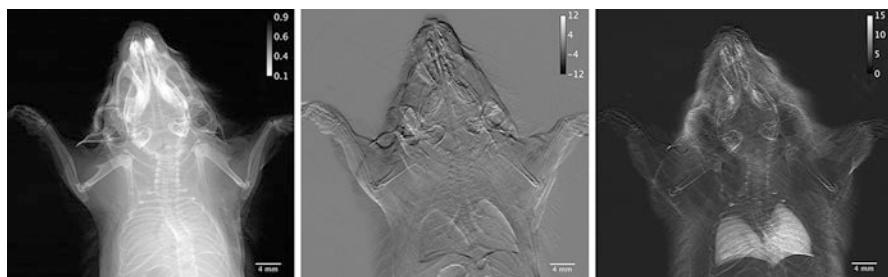
Moving the crystal so that it is positioned at the very tail of the rocking curve for Bragg reflection produces a dark-field image as shown in Fig. 5. In this case only the most strongly refracted or scattered X-rays will be visible in the image. These correspond to very small-scale features due to surface roughness or tiny voids, which may not be resolvable in imaging due to resolution limitations. An example of real ABI data including absorption, refraction, and scattering images is shown in Fig. 7.



**Fig. 5** Sketch of setup for phase-contrast imaging using a crystal analyzer as an angular filter of the transmitted rays. The rocking curve represents the diffraction intensity for the main beam versus crystal angle, the peak being at the Bragg angle



**Fig. 6** Left, a rocking curve sketch indicating the  $H$  and  $L$  positions on the rocking curve where the images  $I_L$  and  $I_H$  are acquired, and right, Chapman et al.'s (1997) "diffraction-enhanced imaging" (DEI) equations for extracting the phase and absorption images from  $I_L$  and  $I_H$



**Fig. 7** ABI images of a mouse showing absorption (left), refraction or vertical phase gradient (center), and scattering (right). (From Arfelli et al. 2018, Creative Commons Attribution 4.0 International License. <http://creativecommons.org/licenses/by/4.0/>)

Suortti et al. (2013) gives a detailed review of the many variations of ABI that have been developed, including more accurate methods for determining phase, absorption, and the dark-field signal. Although medical applications (Connor and Zhong 2014) dominate the research in ABI, there have been some attempts to apply the technique to materials science problems.

### **Applications of ABI**

Lohmann et al. (2002) describe the use of DEI at HASYLAB to examine a number of different specimen types including the internal structure of cultured pearls revealing a number of features not visible using conventional radiography methods and which can assist in identifying pearl origins. Also investigated were fossil- and amber-embedded specimens and the integration of a titanium implant into bone.

Applications relevant to the food industry are demonstrated by ABI studies of the internal structure of walnuts and of cracks, holes, and porosity in wine corks (Donepudi et al. 2010; Rao et al. 2013).

At the interface of materials and medical science, a number of analyzer-based imaging studies have investigated the integration of various types of implants with soft tissue and bone. One of the earliest such studies by Wagner et al. (2006) at HASYLAB investigated the integration of titanium implants into the bone demonstrating the diagnostic value of the method for identifying incomplete integration or loosening. Appel et al. (2016) used refraction contrast to distinguish hydrogel-implanted beads from soft tissue and to follow hydrogel structure and mineralization over time. DEI-CT has also been used to observe biodegradable polymer implants in soft tissue with clear benefits over other imaging techniques for this application (Izadifar et al. 2014).

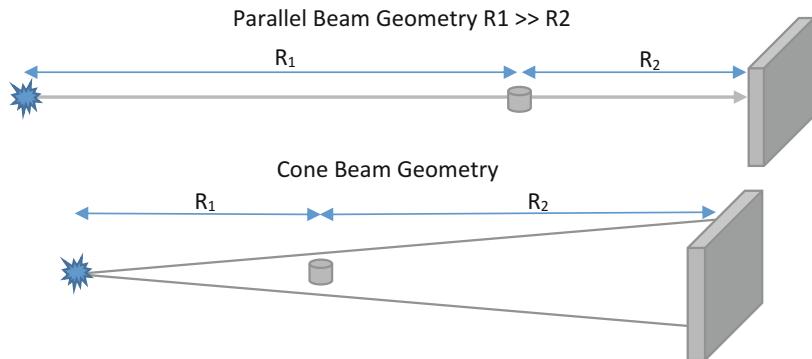
Takeya and colleagues have used DEI-CT to image low-Z materials relevant to energy applications. This includes imaging ice coating on the surface of natural gas clathrate hydrates in order to better understand the conditions in which clathrates can be used to store natural gas and also the observation of water accumulation and ice formation on the carbon paper gas diffusion layers used in fuel cells (Takeya et al. 2012, 2013).

## **Propagation-Based Phase-Contrast Imaging (PB PCI)**

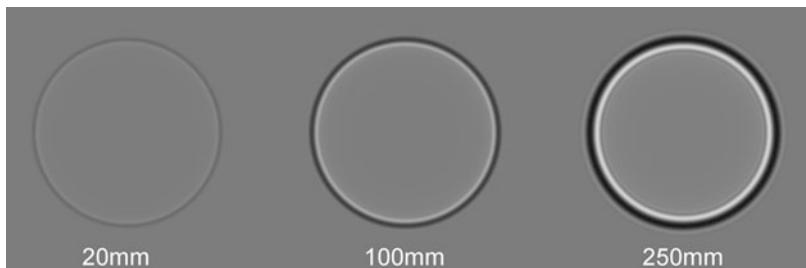
### **PB PCI Fundamentals**

Propagation-based phase-contrast imaging (PB PCI) is the simplest of the phase-contrast imaging modalities to realize experimentally. Like other methods it makes visible the phase shift imposed on an X-ray beam by a sample, but it does this by simply allowing the beam to propagate between the sample and the detector and for this reason is also known as in-line phase contrast. This is analogous to Gabor's development of in-line holography for electron microscopy (Gabor 1948) (Fig. 8).

Propagating X-ray beams, just like their visible light counterpart, will display Fresnel diffraction effects on propagation past a sample or aperture as a result of interference between beams deflected by different amounts as shown in the upper section of Fig. 2. In the case of X-rays, the refractive index is very close to 1, making



**Fig. 8** Imaging geometry of parallel-beam and cone-beam propagation-based phase-contrast imaging



**Fig. 9** Simulated phase-contrast images of an 80  $\mu\text{m}$  diameter polymer sphere imaged with 30 keV parallel-beam X-rays at three different propagation distances. The fringes at the edge of the sphere increase in visibility and width as the propagation distance increases, eventually moving out of the near-field and into the intermediate-field regime in the last image at 250 mm where multiple fringes are beginning to emerge. The sphere would be too weakly absorbing to be visible in an absorption-contrast image at zero propagation distance

these effects much more subtle and difficult to observe without a long propagation distance beyond the sample. Nonetheless, propagation distances from mm to meters (depending on X-ray energy and imaging resolution) are sufficient to observe propagation-based phase contrast for a range of sample sizes at X-ray energies up to 10s of keV. For X-rays the diffraction effects observed are usually in the near-field imaging regime resulting in an edge-enhanced image with characteristic light-dark fringe pairs running along edges and boundaries in the object, as shown in Fig. 9.

PB PCI was first demonstrated by Jacobsen et al. (1990) who utilized soft X-rays from a synchrotron source and the high resolution of a photoresist to create an X-ray Fresnel diffraction image from an object. Synchrotron experiments using parallel-beam hard X-ray imaging in the near-field regime demonstrated the ability to image objects which are transparent to X-rays using phase effects alone (Snigirev et al. 1995). Wilkins et al. (1996) demonstrated in-line phase contrast using polychromatic laboratory sources in the cone-beam geometry.

The fringe contrast for a parallel-beam image will grow stronger with increasing propagation distance ( $R_2$ ) between sample and detector, and the width of the fringes will increase such that width =  $\sqrt{R_2\lambda}$ , where  $\lambda$  is the X-ray wavelength. Having a significant propagation distance between sample and detector makes the PB PCI setup more sensitive to penumbral blurring due to the X-ray source size, which increases with propagation distance such that resolution is limited to  $\sigma_p = \sigma_s R_2/R_1$  where  $\sigma_p$  is the size of the penumbral blurring and  $\sigma_s$  is the X-ray source size. This will reduce fringe visibility significantly for propagation distances beyond which  $\sigma_p > \sqrt{R_2\lambda}$ . Hence, propagation-based phase contrast requires a high spatial coherence of the beam in order to maintain fringe contrast. In practice this means either using synchrotron sources or laboratory X-ray microfocus sources.

Contrast may also be limited by detector resolution which needs to be sufficient to resolve the fringes, since fringe width increases as the square root of the propagation distance detector resolution may place a lower limit on the propagation distance at which phase contrast may be observed.

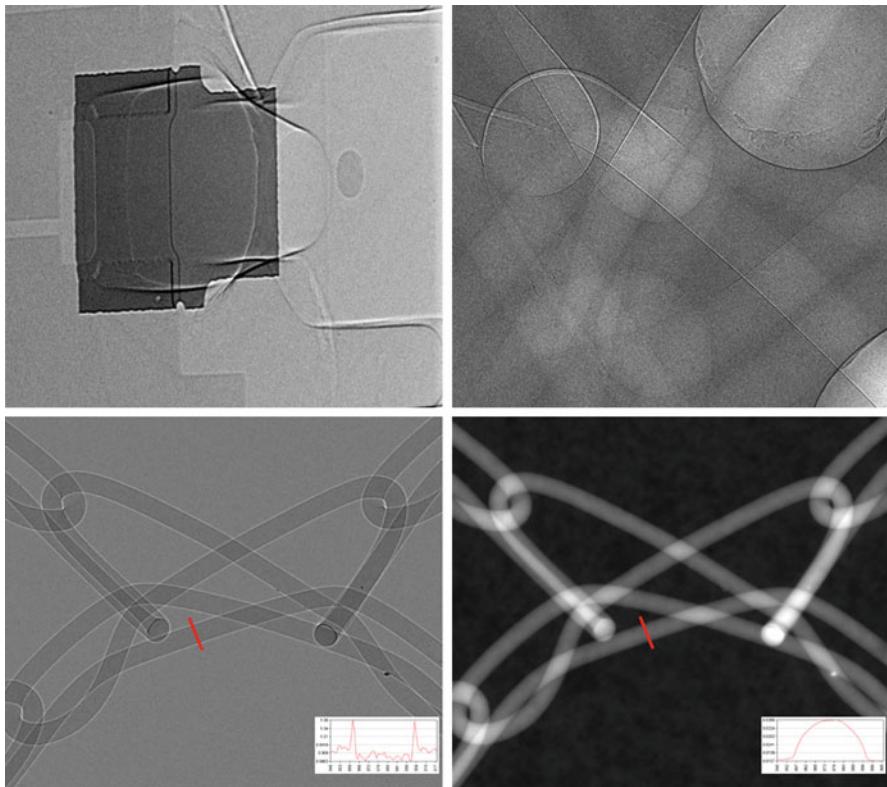
In practice an imaging experiment will typically aim to use sufficient propagation distance to obtain reasonable fringe contrast without hampering resolution via excessively broadening fringes or reducing visibility due to penumbral blurring. In most cases such imaging is done in the near-field, “edge-enhanced” regime in which the fringe size  $\sqrt{R_2\lambda}$  is of the order of the size of the smallest resolvable features, a notable exception being the holographic imaging and holotomography of Cloetens et al. (1999).

For the cone-beam geometry typical of imaging with X-ray microfocus sources, the curvature of the illuminating beam changes the effect of propagation. A cone-beam setup has a magnification of  $M = (R_1 + R_2)/R_1$ , where  $R_1$  and  $R_2$  are the source-to-sample and sample-to-detector distances, respectively. The effective propagation distance for phase contrast in this case is  $R' = R_1 R_2 / (R_1 + R_2)$ .

In this geometry contrast of the fringes increases in proportion to  $R'$  and fringe width (relative to the object plane) in proportion to approximately  $\sqrt{(R'\lambda)}$ , where  $\lambda$  is the average wavelength of the polychromatic beam. The imaging geometry is typically determined by the requirements to magnify a sample of a particular size onto the detector. Achieving good phase contrast under these conditions requires that the penumbral blurring due to the X-ray source size,  $\sigma_s$ , does not blur the phase-contrast fringes on the detector, i.e.,  $M\sqrt{\lambda R'} > \sigma_s R_2/R_1$ , and that the fringes are sufficiently resolved, i.e.,  $M\sqrt{\lambda R'} > \sigma_d$ , for magnification  $M$  and detector resolution  $\sigma_d$ . Detailed analyses of the optimization of phase-contrast imaging systems in terms of contrast and resolution can be found in several papers (Nesterets et al. 2018; Olivo and Speller 2006).

A PB PCI image will contain both absorption and phase contrast so that the intensity in the image does not have a simple relationship to the integral of the attenuation over the path through the sample as is seen for pure absorption-contrast imaging. A raw PB PCI image has instead a strong enhancement of edges and boundaries in the sample which for visual inspection can be a great advantage.

Figure 10 shows examples of PB PCI images in which features of interest are highlighted by PB PCI.



**Fig. 10** PB PC images: (upper left) Ink-jet print head image showing the complementary nature of absorption and phase-contrast image showing different features of the sample; (upper right) polymer composite in which bubbles and delamination between fibers and matrix are highlighted by edge contrast; (lower left) raw PB PCI image of a medical mesh; (lower right) phase-retrieved image of the mesh. Insets show line profiles across a fiber showing how signal to noise is improved by phase retrieval; this produces cleaner data suitable for tomographic reconstruction and segmentation

However, for quantitative analysis phase-retrieval methods are used to recover the phase shift imposed on the beam by the sample. Phase retrieval can reduce noise and produce much cleaner data in a form better suited to tomographic reconstruction and image segmentation as shown in the lower images of Fig. 10.

Many phase-retrieval methods applied to PB PCI data are based on the transport of intensity equation (TIE) which provides a good approximation to PB PCI in the near-field regime (Teague 1983).

$$\nabla_{\perp} \cdot (I(\mathbf{r}_{\perp}, z) \nabla_{\perp} \phi(\mathbf{r}_{\perp}, z)) = - \frac{2\pi}{\lambda} \frac{\partial}{\partial z} I(\mathbf{r}_{\perp}, z).$$

In the case of a pure phase (nonabsorbing) object, if we consider the form of this equation for  $z = 0$ , the image will have no absorption contrast, and therefore its

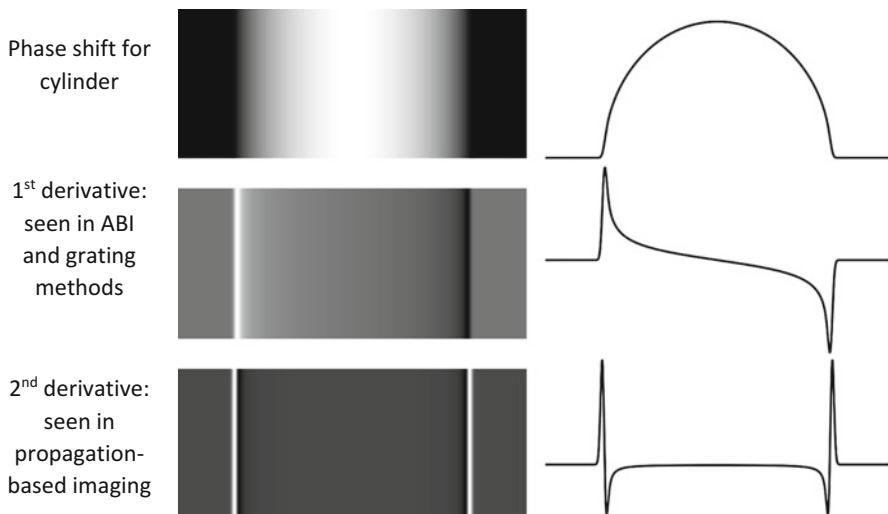
gradient will be zero. In this case the near-field intensity at position  $z$ ,  $I(\mathbf{r}_\perp, z)$ , will be approximated by the equation (Bronnikov 1999)

$$I(\mathbf{r}_\perp, z) = I(\mathbf{r}_\perp, 0) \left( 1 - \frac{\lambda z}{2\pi} \nabla_\perp^2 \phi(\mathbf{r}_\perp, 0) \right)$$

The intensity distribution is thus closely related to the Laplacian (second derivative) of the phase shift imposed on the wave front immediately after passing through the object (see Fig. 11), with the strength of the phase-induced intensity variations increasing in direct proportion to propagation distance. This can be solved to recover the phase shift in the object plane, most conveniently by applying an appropriate filter in the Fourier domain as demonstrated by Nugent et al. (1996) and by Bronnikov (1999) who adapted this approach for use in tomographic reconstruction of phase-contrast data.

Figure 11 shows the phase shift of the X-ray beam that has passed through a cylinder and the corresponding first and second derivatives. The second derivative corresponds to the edge-enhanced Laplacian-type features seen in PB PCI.

For many real-world experimental conditions, the object being imaged will have some X-ray absorption, and modification of the “pure phase” phase-retrieval methods is required. Ideally one would like to independently recover phase and absorption in the image plane. However, with a single image, it is not possible to extract the spatial distribution of two independent variables, and some assumptions must be made.



**Fig. 11** Phase shift profile for a cylinder (top) and its first (middle) and second (bottom) derivatives with corresponding line profiles shown to the right. ABI and grating methods are sensitive to the first derivative of the phase shift, and propagation-based imaging is sensitive to the second derivative

There are a range of different approaches to this problem relying on different assumptions such as a weakly absorbing object or a constant ratio between  $\delta/\beta$ . A detailed comparative review of many of these methods can be found in Burvall et al. (2011). However, the single image algorithm of the type described by Paganin et al. (2002), and its close relatives, is probably the most widely used method due to its computational speed and straightforward experimental requirements. This algorithm assumes a single material (and thus a constant ratio  $\delta/\beta$ ) and solves for sample thickness  $t$ :

$$t(\mathbf{r}_\perp) = -\frac{1}{\mu} \ln \left( \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}(I(\mathbf{r}_\perp, z)/I_0)}{(\alpha|\mathbf{k}|^2 + 1)} \right\} \right), \text{ where } \alpha = z \left( \frac{\delta}{\mu} \right).$$

$\mathcal{F}$  and  $\mathcal{F}^{-1}$  are the Fourier transform and the inverse Fourier transform,  $\mathbf{k}$  is the spatial frequency,  $z$  is the propagation distance,  $(I(\mathbf{r}_\perp, z)/I_0)$  is the normalized image intensity, and  $\mu$  the linear attenuation coefficient.

The contrast transfer function (CTF) approach comprises another class of phase-retrieval methods which extend phase retrieval beyond the near-field regime (Guigay 1977). The potential advantage of CTF methods lies in the fact that they are more sensitive to a broader range of spatial frequencies of phase variation than TIE methods. These include Born or Rytov approximations (Burvall et al. 2011) and the methods of Turner et al. (2004) and Hofmann et al. (2011). “Holotomography” uses images acquired at different propagation distances to enable accurate retrieval of phase information (Cloetens et al. 1999). This produces impressive results but is experimentally challenging due to the requirement for accurate image alignment.

Iterative methods of phase retrieval have a potentially wider field of applicability than the types of approaches outlined above and have been applied in a number of contexts. Myers and colleagues employ prior knowledge about the sample to combine phase-retrieval and binary tomography methods in an iterative scheme which enables quantitative polychromatic phase-contrast tomography of objects composed of up to three materials with greater speed and lower dose than is conventionally required (Myers et al. 2008). Iterative methods have also been applied with dual-energy X-ray imaging for material discrimination (Li et al. 2018) and for improved phase retrieval applicable to monochromatic and polychromatic illumination (Carroll et al. 2017).

## Applications of Propagation-Based Phase-Contrast Imaging and Tomography

The ease with which PB PCI can be implemented means that it has found a wide range of applications, an overview of which is given below. A more detailed review with case studies can be found in Mayo et al. (2012).

### Fracture and Failure

Cracks in metals and alloys have been extensively explored using PB PCI. In situ experimental rigs can be used to apply cyclic or monotonic loading sequences to

samples in the synchrotron beam. Such experiments typically pause to acquire images or tomographic datasets at multiple time-points in the loading sequence to enable the observation of cracking or void formation resulting from cyclic fatigue or static loads. Toda and co-workers have carried out a series of such studies of damage evolution in Al alloys and steel using them to calculate local crack-driving forces and to examine void formation (Toda et al. 2016). PB PC imaging and tomography have also been combined with diffraction-contrast tomography in studies of nickel super alloys (Messe et al. 2014).

Kobayashi and Toda (2007) also investigated cracks in foams of pure Al and Al alloys by combining propagation-based phase-contrast tomography (PB PCT) with other techniques such as local micro-CT and *in situ* methods. They were able to determine that foam morphology and microporosity were the main factors in ductile buckling (pure Al) and brittle fracture (alloy) of cell walls during compressive deformation.

Studies of failure modes in composites include studies of plastic deformation Al/SiC metal-matrix composites (MMCs) (Buffiere et al. 2000), analysis of deformation and damage in T700 carbon fiber/epoxy composites (Bie et al. 2017), and the use of PB PCI to identify delamination in carbon fiber-reinforced polymer composites used in aerospace applications (Vavrik et al. 2015).

Cracking in a variety of other materials has also been analyzed using PB PCI methods. Hu and De Carlo (2008) investigated the formation of cracks in the layered polymer foams that form the insulation on the space shuttle fuel tanks which had been subject to catastrophic failure. Studies have also been made of the failure of biocompatible acrylic cements used in knee and hip procedures (Sinnett-Jones et al. 2005) and of the fracture mechanisms in metallurgical coke, the strength of which is critical to its performance in a blast furnace (Lomas et al. 2017).

### Low-Density Materials and Low-Contrast Boundaries

A series of studies have made use of PB PCI to examine low-Z materials and low-contrast boundaries. Examples include studies of growth defects in silicon carbide (Kohn et al. 2014), visualization of carbon fibers in carbon-carbon (C/C) (Coindreau et al. 2011), and characterization of boundaries between adhesives and wood in wood and bamboo composites (Paris et al. 2015).

Low-Z materials have found potential applications in nuclear technology. PB PCI has been used for imaging of pyrocarbon and silicon carbide coatings for use in fuel pellets for high-temperature reactors (Kashyap et al. 2011) and of inertial confinement capsules used in fusion experiments, (Kozioziemski et al. 2005). PB PCT has been used to quantify the swelling and pore structure in beryllium after 15 years in a research reactor due to the accumulation of helium and tritium in the beryllium matrix (Vladimirov et al. 2015).

X-ray absorbing materials can have low-contrast internal boundaries which benefit from the use of PB PCI. This can be seen in studies of the mechanical behavior of  $\text{Al}_2\text{O}_3/\text{Al}$  metal-matrix composites (Young et al. 2009) and of the ferrite-cementite microstructure in steel (Kostenko et al. 2012).

Wood has a complex hierarchical 3D structure formed largely from low-Z cellulose. PB PCT has been used to extract porosity and microstructural parameters such as

cell wall thickness with an imaging resolution down to 1.5 μm (Trtik et al. 2007). More recently there have been studies of hygroscopic swelling and shrinkage in wood exposed to varying levels of moisture, in particular, investigating the anisotropy, hysteresis, and effect of wood cell size on this process (Rafsanjani et al. 2014). Turning to other botanical applications, Hu et al. (2016) imaged trapped gas films on submerged super hydrophobic leaves and similar bio-inspired surfaces. Dynamic processes in plants have also been observed such as water transport, cavitation, and refilling in the vessels of plants subject to water stress (Xue et al. 2013).

### Batteries and Fuel Cells

A number of studies have demonstrated the utility of phase-contrast methods for the study of components of batteries and fuel cells, many of which are composed of low-Z, low-density materials (Pietsch and Wood 2017). PB PCI has been used to characterize morphology and transport in gas diffusion layers (GDL) and porous graphite electrodes (Becker et al. 2009; Eastwood et al. 2014).

Modes of failure and degradation in batteries and fuels cells are also an active area of research, particularly since failure of these components can be catastrophic. Sun et al. (2017) showed that damage of the three-layer separator due to growth of lithium microstructures during charge and discharge was a significant failure mechanism. In fuel cells Eberhardt et al. (2014) demonstrated that PB PCT was a superior method of determining the distribution of phosphoric acid (PA) in high-temperature polymer electrolyte membrane fuel cells (HT-PEFCs) to gain insight into the mobility and loss of phosphoric acid (PA) which is a factor in loss of performance.

### Soft Matter and Food Science

PB PCI has been applied in a number of applications pertaining to fluids, soft matter, and food science. Studies of fluids include observations of spontaneous meniscus formation in capillaries and of the step-by-step formation of aqueous foams (Andrukh et al. 2014; Garcia-Moreno et al. 2017).

The paper has been imaged using PB PCI from the early days of the technique's development as it provides excellent contrast for low-density cellulose fibers. For instance, it has been applied to the study of 3D structure and transport properties in dry and soaked paper samples (Holmstad et al. 2006). A similar application is the measurement of tensile deformation in mats of electrospun polyester fibers (Maksimcuka et al. 2017).

The potential of PB PCI/PCT for food science research was first noted in Barigou's review of micro-CT for food microstructures (Barigou and Douaire 2013). PB PCT has been used for studies of dough rising and baking (Mayo et al. 2016) and of the effect of thermal cycling (as experienced during transport) on the complex microstructure of ice cream (Guo et al. 2017).

### Paleontology and Cultural Heritage

Paleontology and cultural heritage applications have the common feature that the samples are often extremely rare and precious and often cannot be sectioned or otherwise altered so that nondestructive techniques are essential. PB PCI offers the

possibility of increased contrast for many samples with no more difficulty than using conventional techniques. This has made a significant impact in paleontology in particular with numerous studies of rock fossils (Tafforeau et al. 2006) and fossils embedded in amber (Soriano et al. 2010).

Fossilized ambers have been found containing many kinds of ancient insects and other specimens. A number of PB PCT studies of amber fossils have revealed morphological structures and enabled virtual dissections of sufficient detail to support the taxonomic description of new fossil species (Peris and Hava 2016; Perreau 2012). Fossils imaged and identified in this way include centipedes, beetles, weevils, pseudoscorpions, and conifer tissues, ranging in age from the Cretaceous to the Miocene.

Fossils found in rock are much more strongly X-ray absorbing than those found in amber; nonetheless, phase contrast can considerably improve the visibility of boundaries and small features in such specimens. Such studies have included detailed morphological studies of skeletal and dental morphology (Johanson et al. 2015), PB PCT of fossil plants (Moreau et al. 2014), and high-resolution imaging of Ediacaran fossil embryo-like metazoans (Yin et al. 2016).

Like fossils, PB PCI offers improved contrast and detail for nondestructive studies of rare cultural artifacts such as the Fioravanti et al. (2017) study of the wood types used in historic bows. PB PC laminography is well suited to planar samples and has been used to provide high-resolution 3D maps of the paint stratigraphy in paintings which can be used to understand the evolution of the artwork (Reischig et al. 2013). More recently Mocella et al. (2015) have examined fragile rolled papyri using PB PCT, revealing written letters by virtually “unrolling” the digital representation of the scrolls.

### Medical Scaffolds and Tissue Engineering

Scaffolds, cements, and implants are increasingly used in medical and dental settings for tooth and bone repair and for repair of soft tissue industries. The potential of phase-contrast methods for characterizing these materials has been noted by Appel et al. (2011) and Olubamiji et al. (2014). Applications include studies of bone ingrowth into bone repair materials (Rusu et al. 2014), deformation in dental titanium implants (Zabler et al. 2012), and comparison of different composite resins for tooth repair (Fatima et al. 2016).

### Geosciences

PB PCT has been applied to studies in different areas of geosciences such as analysis of morphological changes due to dehydration in gypsum and using phase-retrieval methods to improve the ability to distinguish different mineral phases in rocks (Arzilli et al. 2016; Xiao et al. 2012). Quantitative PB PCT of quartz-hosted melt inclusions has been used to estimate quartz crystallization times and growth rates to gain insight into magmatic processes (Pamukcu et al. 2013). A more direct approach to observing the behavior of melts has been demonstrated using a specialized sample cell for imaging in liquids and melts at high temperature and high pressure, with the ability to measure viscosity and to distinguish boundaries between immiscible phases via attenuation and phase contrast (Kono et al. 2015).

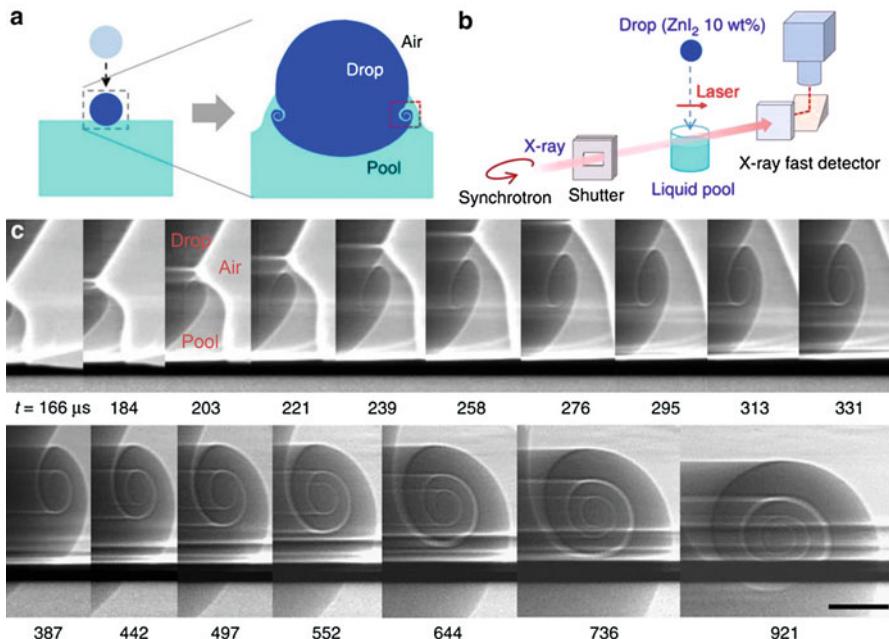
## High-Speed Studies of Dynamic Processes

A series of developments in high-intensity synchrotron sources, high-speed detectors, and related technology have led to a huge growth in the study of dynamic processes and systems using PB PCI methods.

The high-speed study of fracture dynamics has been enabled by the development of synchrotron facilities for in situ rapid PB PCI of samples under loading, most notably at APS which has developed a Hopkinson/Kolsky bar setup on a high-speed imaging beamline (Hudspeth et al. 2013). This kind of apparatus is used to study the dynamic effect of high strain rates on materials including glass (Parab et al. 2014), C/SiC composites (Li et al. 2015), and granular materials (Zbib et al. 2015).

Fluid dynamics research is another major area of activity in high-speed PB PCI. PB PCI is very sensitive to surfaces parallel to the propagating beam which gives rise to fringes in the image, such that PB PCI images of simple structures like falling drops give an excellent view of the cross section of these surfaces. Lee and colleagues have used these images to examine the dynamics of drops on impact, including observation of the formation of vortices (Lee et al. 2015) (Fig. 12).

Other fundamental fluid dynamics studies have included analysis of cavitating flows in water and flow paths of a particle suspension flowing though porous media (Blankenburg et al. 2017; Khelifa et al. 2017).



**Fig. 12** Dynamics of the edge of a falling drop of ethanol (including contrast agent) impacting a pool of pure ethanol captured using high-speed PB PCI. (Figure from Lee et al. 2015, Creative Commons Attribution 4.0 International License. <http://creativecommons.org/licenses/by/4.0/>)

More industrially focussed applications of high-speed PC PCI include investigations of the dynamics of the electro-spinning process (Greenfeld et al. 2012) and of the solid-liquid interface and porosity formation in laser welding (Miyagi et al. 2017). Numerous studies of flow and sprays from nozzles have been carried out in the contexts of fuel injectors for internal combustion engines (Duke et al. 2015; Moon et al. 2015) and pressurized metered-dose inhalers in order to determine the optimum conditions for effective respirable drug delivery (Mason-Smith et al. 2017).

### Laboratory-Based In-Line Phase-Contrast Imaging for Material Characterization

Laboratory-based in-line phase-contrast imaging has run in parallel to developments at synchrotrons and as outlined above typically makes use of microfocus X-ray sources and increasingly turnkey micro-CT systems available from a range of manufacturers. Despite the lower flux and consequent longer data collection times, the accessibility of lab sources in comparison to synchrotrons makes them attractive for many applications.

When using a microfocus X-ray source for high-resolution imaging at modest X-ray energies, phase contrast is not only achievable but it is almost unavoidable. Hence a full appreciation of PB methods and appropriate phase-retrieval techniques is highly desirable to achieve optimum results in imaging and micro-CT.

Gureyev et al. (1999) first applied lab-based PB PCI technique to 2D imaging of paper and wood demonstrating excellent edge contrast to observe their microstructure. An early assessment of the potential of the technique for nondestructive materials evaluation is given by Zoofan et al. (2006), and a more recent review of laboratory applications is found in Mayo et al. (2012).

Laboratory-based PB PCI of low-density materials was demonstrated in a study of interfaces in graded aerogels with densities in the range of 20–200 mg/cm<sup>3</sup> and in a series of studies of surrogates for inertial confinement targets for nuclear fusion applications (Gui et al. 2011; Wang et al. 2014). Kastner et al. (2012) demonstrated the value of PB PCI methods in imaging polymer composites and alloys. Studies of composites include PB PCT of carbon/carbon composites (Fu and Liu 2016) and visualization of delamination in a carbon fiber-reinforced plastic (CFRP)-layered composite material (Vavrik et al. 2015).

Studies of natural materials include foodstuffs including investigations of Brazil nuts and a quantitative analysis of the morphology of wheat-corn-based extrudates (Franco et al. 2015; Pitts et al. 2016). Biofilms are another low-Z biomaterial which has been imaged using a combination of a nondestructive contrast agent with PB PCI (Carrel et al. 2017).

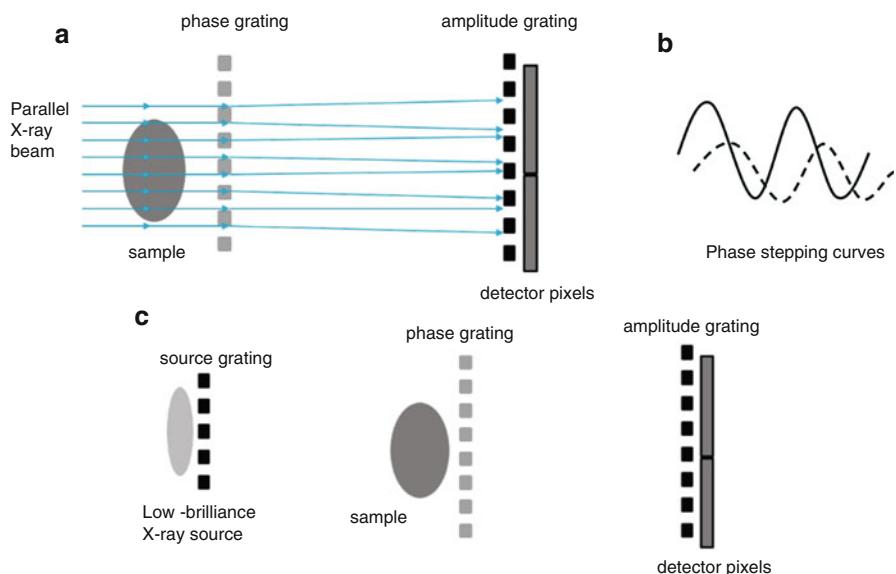
### Grating-Based Phase-Contrast Imaging (GB PCI)

#### GB PCI Fundamentals

Grating-based phase-contrast imaging (GB PCI) exploits periodic structures to perturb and analyze the X-ray beam. A typical experimental setup is depicted in

Fig. 13: a well-collimated X-ray beam is modulated by a phase grating and analyzed by an amplitude grating placed downstream, before the detector pixels. The distance between the two gratings is chosen to maximize the contrast of the interference fringes, in relation to the Talbot self-imaging of periodic structures (Talbot 1836). When the amplitude grating is laterally shifted, the intensity recorded by the detector follows a periodic modulation. This can be used as the basis to separate the contributions from sample absorption, phase shift, and scattering. Typical phase-stepping curves are sketched in Fig. 13b, where the solid line indicates the intensity modulation in the absence of the sample in the beam, while the dashed curve is the modulation when the sample is present. Three effects can qualitatively be identified, an overall reduction of the intensity reaching the pixel, linked to the sample absorption; a shift of the modulation pattern, linked to the gradient of the phase shift imposed by sample; and a reduction in fringe visibility, linked to scattering.

Early proposals and investigation of these effects in the hard X-ray regime date back to the 1990s (Clauser and Reinsch 1992; Cloetens et al. 1997) with imaging experiments following soon afterward (David et al. 2002; Momose 2005; Momose et al. 2003; Weitkamp et al. 2005). The intensity  $I(x)$ , recorded by a detector pixel as a function of the amplitude grating displacement, can be quantitatively expressed as



**Fig. 13** (a) Sketch of a grating-based X-ray phase-contrast imaging setup. High-brilliance radiation, travelling from the left- to the right-hand side, traverses the sample and passes through a phase grating. The amplitude grating, placed downstream and before detection, modulates the intensity recorded by the detector pixels. (b) Illustrative phase-stepping curves with (dashed) and without (solid) the sample in the X-ray beam: the intensity recorded by a pixel is modulated as a function of the lateral translation of the amplitude grating. (c) Three-grating setup using low-brilliance X-ray sources: an additional amplitude grating is employed to structure the extended source into an array of spatially coherent but mutually incoherent secondary sources

$$I(x) = \sum_i a_i \cos\left(\frac{2\pi i x}{p_2} + \phi_i\right) \approx a_0 + a_1 \cos\left(\frac{2\pi x}{p_2} + \phi_1\right)$$

where  $a_i$  are the amplitude coefficients,  $\phi_i$  is the phase coefficients, and  $p_2$  is the period of the grating. The images of the sample are retrieved by comparing the phase-stepping curves recorded with and without the sample in the beam, at each detector pixel position.

The sample transmission image is given by

$$T = a_0^s/a_0$$

where the superscript (  $s$  ) indicates the presence of the sample in the beam. The differential phase-contrast image is obtained by taking the difference

$$\nabla_x \Phi = \frac{p_2}{\lambda d} (\phi_1^s - \phi_1)$$

of the measurements taken with and without the sample in the beam. The radiation wavelength,  $\lambda$ , and the distance between the two gratings,  $d$ , are taken into account to retrieve the component of the gradient of the phase shift imparted by the object to the incoming wave front.

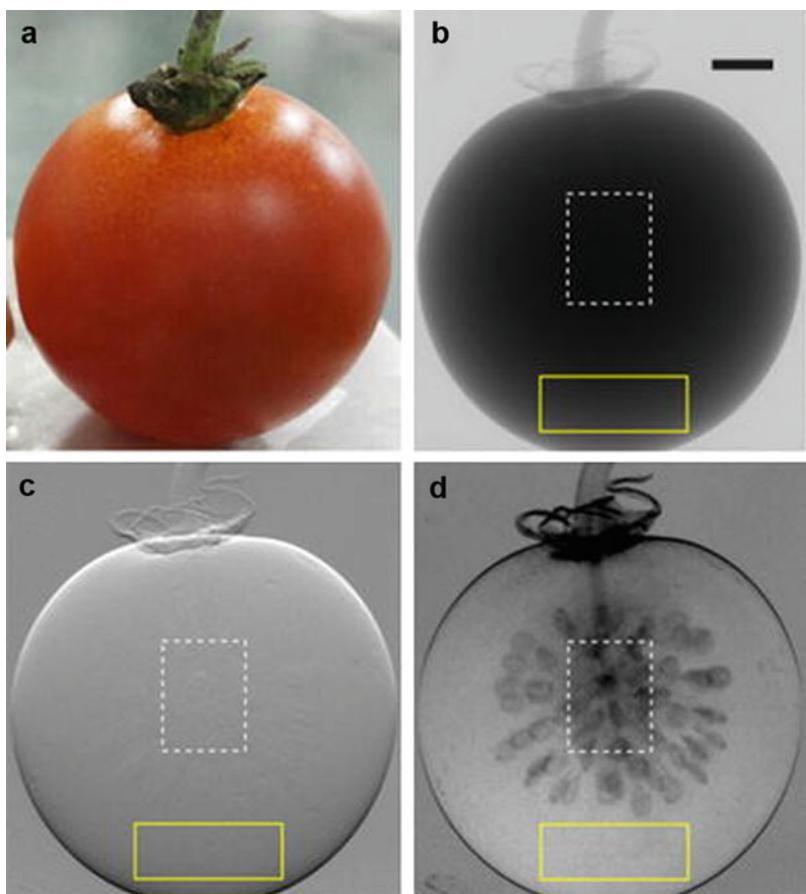
Dark-field contrast is linked to the local scattering power of the sample and is evaluated by looking at how it affects fringe visibility. By defining the fringe visibility as the ratio between the first two amplitude coefficients, the sample's contribution to the decrease of such visibility can be evaluated by calculating the ratio

$$V = \frac{a_1^s a_0}{a_0^s a_1}.$$

This quantity remains unchanged ( $V = 1$ ) for samples with negligible scattering power, e.g., highly homogenous materials, while a sizeable reduction in visibility can be measured ( $V < 1$ ) when strong density fluctuations are present at the subpixel scale lengths.

A major breakthrough for this imaging technique was the development of the Talbot-Lau-type hard-X-ray interferometer that made GB PCI phase and dark-field imaging compatible with low-brilliance X-ray sources (Pfeiffer et al. 2006, 2008). A schematic representation of this type of interferometer is given in Fig. 13, where a third amplitude grating is used to create an array of individually coherent, however, mutually incoherent, sources.

The Talbot-Lau laboratory setup makes GB PCI much more accessible and enables multimodal imaging analogous to that seen in ABI readily achievable in the lab (see Fig. 14). This has opened up the technique to a wider range of applications, and one of particular interest is directional dark-field imaging. GB PCI using linear gratings is only sensitive to the component of phase shift and scattering in the direction perpendicular to the grating lines. This means that it is possible to measure anisotropy in scattering from highly anisotropic textures in a sample by measuring the scattering signal as a function of the relative rotation of the grating (around the optic axis) with

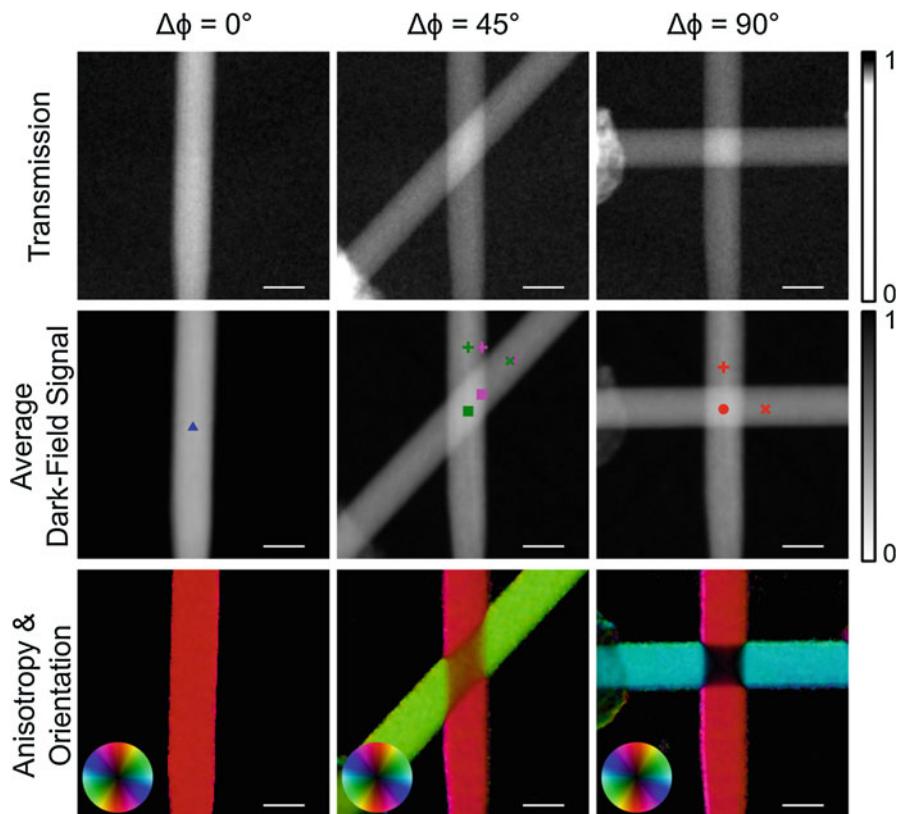


**Fig. 14** GB PCI images of (a) a cherry tomato showing (b) absorption, (c) refraction or phase shift in the vertical direction, and (d) intensity loss due to scattering from Wang et al. (2017) (Copyright Springer). The squares indicate regions with very different changes in intensity between the different imaging modes

respect to the sample (Fig. 15). Some practical applications of directional dark-field imaging in analysis of textured materials are discussed in the following section.

In addition to the grating setups described above, a range of other grating interferometry schemes have been investigated. For example, two-dimensional gratings have been used for extraction of directional phase and scattering information without rotating the grating (Zanette et al. 2010), and sampling gratings with narrower openings than conventional absorption gratings have been used to increase phase sensitivity and data acquisition speed (Du et al. 2015).

A range of “single-shot” grating methods which avoid the need to acquire multiple images to extract multimodal data are also under active development. These are discussed further in the “Emerging Techniques” section.



**Fig. 15** GB PCI images of wooden skewers in different orientations showing the transmission images (top row), the dark-field images (middle row), and the scattering anisotropy and direction obtained using directional dark-field imaging (bottom row). (From Malecki et al. 2013, <http://journals.plos.org/plosone/article/figure?id=10.1371/journal.pone.0061268.g003>, under Creative Commons Attribution License)

### Applications of Grating-Based Phase-Contrast Imaging and Tomography

Grating-based phase-contrast imaging and tomography (GB PCI, GB PCT) have experienced an enormous growth in recent years as these relatively new methods have become mature and better established. A great deal of this work has been focussed on bringing these methods into the clinic for medical applications (Gkoumas et al. 2016; Olubamiji et al. 2014); however, there is also a significant body of research exploring the potential of these techniques for nondestructive imaging in nonmedical fields. Particularly notable is the large number of successful studies carried out in a lab setting, thanks to the Talbot-Lau approach, and the value of dark-field imaging in detecting features below the resolution limit of the imaging system.

### Wave Front Sensing for Optical Metrology

One of the earlier applications of GB PCI was the characterization of X-ray optical components through wave front sensing methods. Visible Moire fringes generated by the gratings are one way of assessing the effect of the optic on the X-ray wave front. The more typical stepping approach has been applied to assessing X-ray monochromator distortion due to heating (Rutishauser et al. 2013).

### Materials

GB PCI and PCT have had a number of successes in materials science applications, with dark-field imaging playing an important role in many cases. Sarapata et al. (2015) used synchrotron Talbot GB PCT to observe the structure of concrete in fine detail, noting the value of dark-field imaging for detecting micro-cracks. The setting of cement was observed in a time-resolved lab-based study by Prade et al. (2015), who observed the dissolution of cement particles over time through the corresponding reduction in the dark-field signal and finding good agreement with ultrasound methods.

GB PCI is sensitive to features in the direction perpendicular to the grating lines; in the case of an anisotropic structure, valuable information about texture can be obtained by completing directional GB PCT scans with the sample in two orthogonal orientations. Directional dark-field imaging was used by Revol et al. (2013) to characterize the fiber structure in carbon fiber-reinforced polymers (CFRP), by Reza et al. (2014) to evaluate paperboard quality, and by Malecki et al. (2014) in a study of the mechanical properties of wood. The EVITA project is a feasibility study for using GB PCI for NDI of larger samples of aerospace CFRPs and has combined GB PCI with image processing techniques to detect directional defects (Gresil et al. 2017).

Jerjen et al. (2013) used GB dark-field imaging in a lab Talbot-Lau setup to detect stress whitening in plastics. They found this could be assessed more rapidly and for larger sample areas than is possible using micro-CT, with potential for screening plastic parts on the production line. Another polymer study by Wu et al. (2017) used a Talbot-Lau grating setup with a powerful synchrotron white beam to do time-resolved GB PCT of phase separation in a melting polymer blend.

### Industrial Inspection and Security Scanning

Two applications which focus on rapid assessment of samples are industrial inspection and security scanning. Uehara et al. (2013) trialed a Talbot-Lau GB PCI system employing a unique multiline X-ray source instead of a first grating for inspection of integrated circuit (IC) package and power modules. Miller et al. (2013) explored the potential of GB PCI lab-based systems for security scanning, specifically the detection of powdered explosives, observing the usefulness of the dark-field mode to detect these materials.

### Tissue Engineering and Related Products

Like PB PCI, there have been a number of studies using GB PCI to address the characterization of scaffolds and similar tissue engineering products. Examples

include the Cedola et al. (2014) synchrotron study of bone growth into a porous ceramic scaffold for bone repair after 48 days in samples extracted from mice and a study by Takashima et al. (2015) observing degradation of implanted biodegradable scaffolds in spinal cords extracted from rats 7 and 21 days after implantation.

### Food Science

Characterization of food products is important for quality control and also for fundamental studies in food science to improve products and processing. Lab-based GB PCI studies of food products have clearly demonstrated excellent visibility of internal features such as seeds in fruit (Wang et al. 2017), in which the dark-field scattering signal was of particular importance to observe the features of interest. An important quality control issue for food is the identification of foreign objects contaminating food products. Einarsdottir et al. (2016) have tested the sensitivity of GB PCI in the detection of different types and sizes of foreign objects contaminating a range of food products, demonstrating the superior sensitivity over conventional X-ray scanning, particularly for organic materials.

Miklos and colleagues have used GB PCT to characterize meat products, finding that the different imaging modes offer superior discrimination of different phases including protein, fat, and connective tissue. They compared meat before and after cooking and discussed potential for online screening (Miklos et al. 2016).

### Cultural Heritage

The potential of GB PCI and PCT for examining precious cultural heritage objects is illuminated by two recent studies. Altapova et al. (2012) used synchrotron GB phase-contrast laminography to investigate degradation processes in ancient parchments by comparing fresh parchment with a degraded fragment of Dead Sea parchment, which showed a much stronger dark-field signal indicating deterioration of collagen and formation of micro-cracks. Albertin et al. (2015) conducted a comparison of different synchrotron techniques for the study of historic documents, finding in the case of GB PCI that the dark-field signal could reveal written characters even when the attenuation contrast was too weak to do so.

## Emerging Techniques

### Tracking-Based Phase-Contrast Techniques

Tracking-based phase-contrast encompasses a number of techniques which all utilize highly structured illumination created using a grating or mask and imaged with a detector placed at a suitable distance. The presence of a sample introduced into the beam path modifies the image of the mask, and by analyzing the changes in the intensity, visibility and position of the features in the image, the absorption, refraction, and dark-field images of the sample can be extracted. Like ABI and GB PCI, tracking methods are sensitive to the first derivative of the phase and so have a better sensitivity than PB PCI to low spatial frequency variations in the phase shift imposed by the sample.

The mask used to structure the illumination can be a regular periodic grating, such as a line or mesh pattern (Morgan et al. 2016) or a random mask, like sandpaper (Berujon et al. 2012). The typical feature size of the mask and the propagation distance between the mask and detector are selected to maximize contrast of fine-scale features in the mask image. For a periodic grating, this will usually correspond to the Talbot distances. For a random mask like sandpaper, the distance is less critical; nonetheless, for a given typical feature size (for instance, a particular grade of sandpaper), there will be a range of distances that maximize contrast.

In order to obtain high-resolution results, stepping techniques similar to those used with gratings can be employed (Zdora et al. 2018). However, a major benefit of tracking methods over GB PCI or ABI methods is that a single image can be used for the analysis, albeit at the cost of reduced resolution. Such “single-shot” methods are more readily applicable to the imaging of dynamic systems.

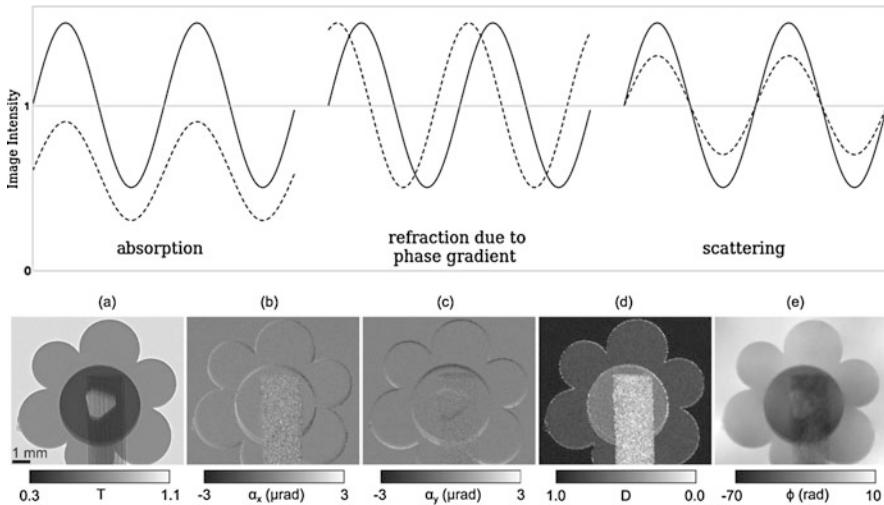
Experiments with regular gratings and random masks have for the most part used one of the two approaches for the analysis. The first is the Fourier-based approach which is applicable to masks of a uniform period across the field of view (Wen et al. 2010). In this method a Fourier transform is taken of the image, and a subset of the transform centered on the peak corresponding to the grating period, calibrated to the sample-free equivalent, is inverse-transformed to obtain the refraction image. For a 2D grating, the refraction images in the x and y directions can be obtained in this way. Scattering images can also be extracted using this method.

The other method, which can be applied to both gratings or a random mask illumination, examines the localized differences between the images with and without the sample (Morgan et al. 2016). The mask image will affect the three ways by the sample: (1) the intensity will be reduced due to absorption, (2) the position of mask features will be shifted due to refraction, and (3) the visibility of mask features relative to local intensity will be reduced due to scattering. From each of these, the sample’s absorption, refraction (both magnitude and direction), and dark-field images, respectively, can be extracted (Zanette et al. 2014). Figure 16 (upper section) shows how the image intensity profile due to a mask will be affected by absorption, refraction, and scattering. In practice all these effects are present simultaneously but can be separately extracted to obtain the multimodal images shown in the lower part of the figure.

This localized comparison between the images with and without the sample is carried out for a small window around each image pixel, the size of which will determine the resolution of the output images. Considering the local displacement of the mask image due to refraction, for an x component of displacement  $S_x$ , the phase gradient along x for the window region is given by

$$\frac{\partial\phi}{\partial x} = \frac{2\pi}{\lambda} \tan^{-1} \left( \frac{S_x}{D} \right)$$

and similarly for the y direction, where  $D$  is the distance between the mask and the detector and  $\lambda$  is the X-ray wavelength. Morgan et al. (2011) maximize a local cross-correlation for this window between the images with and without the sample over a



**Fig. 16** Upper: sketch of how a line profile through a reference mask image (solid line) is affected by absorption, refraction, and scattering (dashed lines). Lower: speckle-based multimodal images of a plastic flower on a wooden rod, acquired using a “single-shot.” (a) transmission, (b) refraction along x, (c) refraction along y, (d) dark-field, and (e) phase shift (determined by integrating the refraction signal). (Reprinted with permission from Zanette et al. (2014). Copyright (2014) by the American Physical Society)

small range of shifts between the two to find the local image displacement, using subpixel interpolation for accuracy. The resultant gradient images are then integrated to construct the phase-shift image.

Zanette et al. (2014) describe a similar approach where the effect of transmission and scattering is built into a model for the local effect of the sample on the mask image. This is used to generate a cost function which is minimized to find local image displacement and thus to generate phase, transmission, and scattering images. The effects of polychromicity can also be accounted for to produce more accurate results for laboratory X-ray sources (Vittoria et al. 2015). This latter approach can be successfully applied to standard (non-microfocus) lab X-ray sources thanks to the characteristics of the mask used (Vittoria et al. 2017).

Tracking methods are relatively new but have already found applications in the medical (Morgan et al. 2014; Wen et al. 2009) and metrology fields (Berujon et al. 2014) including successful demonstrations of directional dark-field imaging (Kagias et al. 2016).

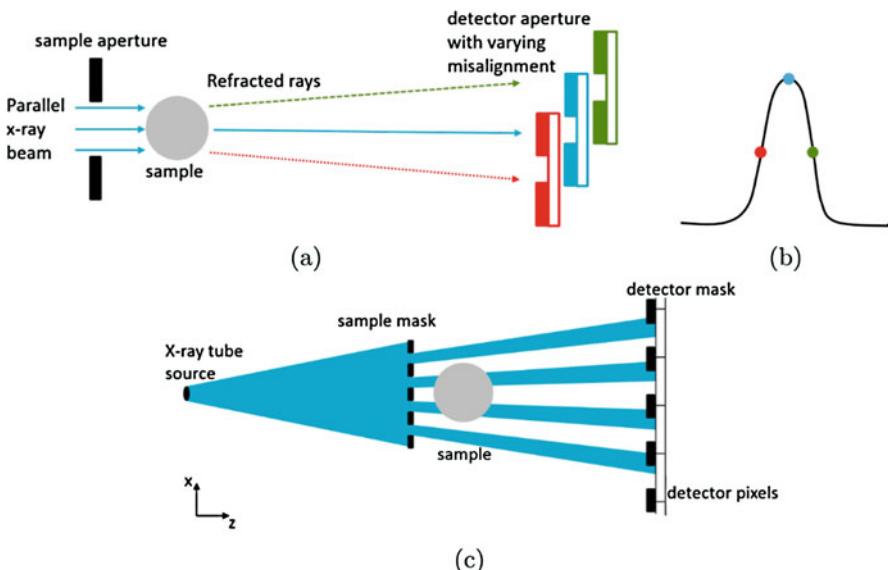
### Edge Illumination

Edge illumination (Olivo et al. 2001) (EI) was firstly developed as a synchrotron-based X-ray phase-contrast imaging method in the late 1990s. It takes inspiration from crystal analyzer techniques and achieves fine angular selection by means of narrow absorbing apertures. A typical synchrotron setup is sketched in Fig. 17:

a well-collimated and monochromatic synchrotron radiation beam is shaped down to a narrow blade of radiation by an aperture placed before the sample.

After propagating through the sample, the beam is analyzed by a second aperture in close vicinity of the detector. The function describing how the detected intensity changes as a function of the relative displacement between the sample and the detector aperture is usually called illumination function. Minimum intensity is detected when the two apertures are misaligned, while the maximum occurs when they are perfectly aligned. This function is key in the characterization of an EI imaging system and in understanding its working principles. By referring to Fig. 17, when the system is aligned in such a way that half of the X-ray beamlet impinges on the absorbing edge of the detector aperture, the working points on the *IF* are halfway down the slopes of the *IF* (red and green dots in panel (b)). A shift of the X-ray beam, as induced, for example, by a phase gradient in the path, is transformed into detected intensity modulation by the analyzer.

The translation of these concepts to a laboratory system (Olivo and Speller 2007) is typically achieved by replicating the single-beamlet arrangement (Fig. 17a) for as many row or columns of detector pixels as possible are needed in order to cover the desired field of view. A sketch illustrating a laboratory implementation is shown in Fig. 17c. This approach proved to be robust against thermal and mechanical stresses and fully exploits the wide energy bandwidth produced by an X-ray tube.



**Fig. 17** (a) Typical edge illumination setup for an experiment with synchrotron radiation. A narrow aperture strongly structures the radiation field before it reaches the sample, and a second aperture serves as an analyzer before detection. (b) Example illumination function. (c) Typical laboratory implementation: the polychromatic and divergent beam produced by an X-ray tube is modulated by a series of apertures preceding the sample, and it is analyzed by another set of apertures that are harmonically matched to the detector pixels

The geometrical optics approximation is often accurate in predicting the system behavior for typical laboratory systems. In these cases, the *IF* can be modelled as the convolution of the detector aperture, the sample aperture, and the source intensity distribution (Endrizzi et al. 2014). Consider a single aperture of the system in Fig. 17c, the intensity distribution projected at the detector plane is given by

$$I_1(x) = \int dx' S\left(\frac{x}{G-1}\right) \text{rect}\left(\frac{x}{a_1 G} - x'\right)$$

where  $S(x)$  represents the intensity distribution at the source plane,  $\text{rect}(x)$  the rectangular function of value 1 for  $-1/2 < x < 1/2$ , and zero elsewhere. The geometrical magnification factor is  $G$ , given by the ratio between the source-to-detector and the source-to-sample distances, and  $a_1$  is the aperture width in the sample mask. The total intensity recorded by a pixel as the detector aperture is shifted along the  $x$  axis is given by

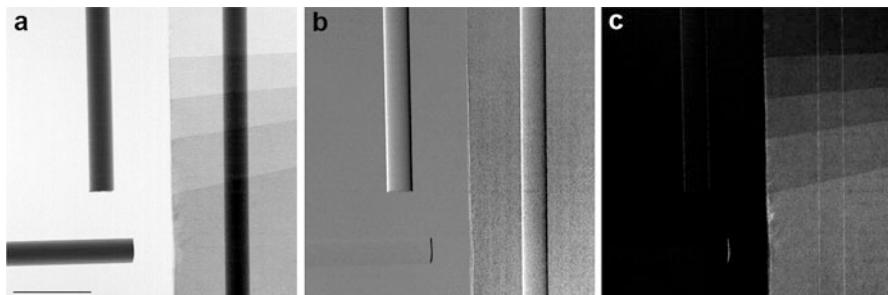
$$\begin{aligned} IF(x) &= \int dx' I_1(x) \text{rect}\left(\frac{x}{a_2} - x'\right) \\ IF(x) &= \iint dx'' dx' S\left(\frac{x}{G-1}\right) \text{rect}\left(\frac{x}{a_1 G} - x'\right) \text{rect}\left(\frac{x}{a_2} - x''\right) \end{aligned}$$

where  $a_2$  is the aperture width in the detector mask. When a sample is present in the X-ray beam path, the intensity reaching the detector is modulated in three ways. Sample transmission indicates the amount of radiation which is removed from the beam and is given by the ratio between the emerging and impinging X-ray intensities (attenuation contrast). Refraction changes the overall direction of propagation of the X-rays, resulting in a global shift of the *IF* (differential phase contrast). Scattering increases the angular spread of the beam emerging from the sample, which is reflected by an increased width of the *IF* (dark-field contrast). The presence of the sample in an EI system can be modelled as a convolution between the *IF* and an object function  $O$  (Endrizzi et al. 2014, 2015a)

$$I(x) = \int dx' IF(x) O(x - x')$$

that summarizes the modifications imposed by the sample on the *IF*. More formally, the sample transmission is given by  $t = \int dx O(x)$ , the sample refraction by  $\Delta x = (1/t) \int dx O(x)x$ , and the sample scattering by  $w = (1/t) \int dx O(x) (x - \Delta x)^2$ . In an EI system, these equations hold independently on a pixel-by-pixel basis.

EI enables phase and dark-field retrieval under practically negligible coherence conditions (Endrizzi et al. 2014; Munro et al. 2012). Phase and absorption can be retrieved by using just two intensity projections, acquired on opposite slopes of the *IF*, and dark-field images can be retrieved by recording a third intensity projection. Analytical solutions (Endrizzi et al. 2014; Endrizzi and Olivo 2014; Munro et al.



**Fig. 18** Demonstrative phantom for an EI system: (a) transmission, (b) refraction, and (c) dark-field images of a phantom made of three acrylic cylinders and a paper step wedge. The scale bar in the bottom-left corner of panel (a) is 1 cm. (Reproduced from “Hard X-ray dark-field imaging with incoherent sample illumination” Appl. Phys. Lett. 104, 024106 (2014); <https://doi.org/10.1063/1.4861855>, with the permission of AIP Publishing)

2012) are available for the cases where assumptions on either the system’s characteristics or the sample allow for a simplified description of the image formation process. Numerical solutions (Diemoz et al. 2013; Endrizzi et al. 2015a; Modregger et al. 2017) provide the retrieval of the sample’s properties without requirement of any a priori knowledge and take into account also the local imperfections of the imaging system (as, e.g., mask defects, nonuniformity of the radiation field, and detector response inhomogeneity). Single-shot approaches have been developed by assuming quasi-homogeneous sample and retrieving its thickness or projected electron density (Diemoz et al. 2015) and by working under phase-object approximation for differential phase-contrast and dark-field imaging (Endrizzi et al. 2018).

Figure 18 shows the three-channel representation of a demonstrative phantom to illustrate the basic characteristics of the images produced by an EI system. The phantom was built by using three acrylic cylinders and a paper step wedge.

The panels show the phantom (a) transmission, (b) refraction, and (c) dark-field contrast channels. The differential nature of the refraction image can easily be seen by looking at how the acrylic cylinders change appearance depending on their spatial orientation relative to the system apertures (running along the vertical direction). While the three cylinders appear with the same contrast in the transmission image, the horizontal one almost disappears in the refraction image. The dark-field contrast, captured in the (c) panel, is linked to the subpixel inhomogeneity of the phantom: the acrylic material vanishes, while paper exhibits clear scattering properties. Both materials give rise to contrast in the transmission image, and by looking at the dark-field image, they can be differentiated based on their microstructure.

Due to the one-dimensional shape of the structuring and analyzing elements, sensitivity in the direction parallel to the aperture is lost. Simultaneous two-dimensional sensitivity can be restored by using sample and detector masks that are patterned in two dimensions (Kallon et al. 2015; Olivo et al. 2009).

EI computed tomography (CT) has been implemented both by using monochromatic synchrotron radiation (Hagen et al. 2014a) and laboratory X-ray tubes (Hagen

et al. 2014b), allowing for low-dose nondestructive inspection of three-dimensional structures. Further developments include single-shot reconstruction algorithms (Diemoz et al. 2017) for fast data acquisition of homogeneous and multi-material samples (Zamir et al. 2017).

Applications of EI imaging span a variety of fields, encompassing medicine, materials science, and security. Examples include low-dose phase-contrast mammography with conventional X-ray tube and detector (Olivo et al. 2013), cartilage imaging (Marenzana et al. 2014), tissue engineering (Hagen et al. 2015), lung imaging (Modregger et al. 2016), composite materials inspection (Endrizzi et al. 2015b), and a large field-of-view high-energy scanning system (Astolfo et al. 2017).

---

## Summary

Phase-contrast imaging has become a widely used and mature technique in the last 20 years. Propagation-based phase contrast is a very accessible method which can be implemented in the lab using standard microfocus X-ray sources and equipment, primarily by utilizing geometry and mathematics to improve image quality. Grating-based phase-contrast requires specialized hardware and multi-step imaging procedures but can be used with non-microfocus high-power X-ray sources and opens up additional imaging modes, such as dark-field imaging, which can shed light on cracks, porosity, and texture on length scales below the inherent resolution of the system.

Emerging tracking and edge illumination methods enable multimodal imaging similar to GB PCI but with simpler hardware and the ability to use single-shot imaging. Research is very active in this area and will likely give rise to faster, simpler, and better quality laboratory-based phase-contrast imaging methods in years to come.

The continued development of synchrotron sources and the emergence of X-ray-free electron lasers (XFEL) will continue to drive the improvements of existing methods such as ABI (Arfelli et al. 2018) and ultrahigh-speed PB PCI (Schropp et al. 2015), as well as the development of new techniques like X-ray ptychography (Pfeiffer 2018) which offers an unrivalled combination of high-resolution and large field of view.

Phase-contrast imaging in all its forms has already made a significant impact in NDI, and its ongoing development will ensure it continues to do so for the foreseeable future.

---

## References

- Albertin F, Astolfo A, Stampanoni M, Peccenini E, Hwu Y, Kaplan F, Margaritondo G (2015) Ancient administrative handwritten documents: x-ray analysis and imaging. *J Synchrotron Radiat* 22:446–451. <https://doi.org/10.1107/s1600577515000314>

- Altapova V, Helfen L, Myagotin A, Hanschke D, Moosmann J, Gunneweg J, Baumbach T (2012) Phase contrast laminography based on Talbot interferometry. *Opt Express* 20:6496–6508. <https://doi.org/10.1364/oe.20.006496>
- Ando M, Hosoya S (1972) An attempt at x-ray phase-contrast microscopy. In: Proceedings of the 6th international conference on x-ray optics and microanalysis. University of Tokyo Press, Tokyo, pp 63–68
- Andrukha T, Monaenkova D, Rubin B, Lee WK, Kornev KG (2014) Meniscus formation in a capillary and the role of contact line friction. *Soft Matter* 10:609–615. <https://doi.org/10.1039/c3sm52164h>
- Appel A, Anastasio MA, Brey EM (2011) Potential for imaging engineered tissues with x-ray phase contrast. *Tissue Eng Part B Rev* 17:321–330. <https://doi.org/10.1089/ten.teb.2011.0230>
- Appel AA, Ibarra V, Somo SI, Larson JC, Garson AB, Guan HF, McQuilling JP, Zhong Z, Anastasio MA, Opara EC, Brey EM (2016) Imaging of hydrogel microsphere structure and foreign body response based on endogenous x-ray phase contrast. *Tissue Eng Part C Methods* 22:1038–1048. <https://doi.org/10.1089/ten.tec.2016.0253>
- Arfelli F, Astolfo A, Rigon L, Menk RH (2018) A Gaussian extension for diffraction enhanced imaging. *Sci Rep* 8:362. <https://doi.org/10.1038/s41598-017-18367-x>
- Arzilli F, Polacci M, Landi P, Giordano D, Baker DR, Mancini L (2016) A novel protocol for resolving feldspar crystals in synchrotron x-ray microtomographic images of crystallized natural magmas and synthetic analogs. *Am Mineral* 101:2301–2311. <https://doi.org/10.2138/am-2016-5788>
- Astolfo A, Endrizzi M, Vittoria FA, Diemoz PC, Price B, Haig I, Olivo A (2017) Large field of view, fast and low dose multimodal phase-contrast imaging at high x-ray energy. *Sci Rep* 7:2187
- Barigou M, Douaire M (2013) X-ray micro-computed tomography for resolving food microstructures. In: Morris VJ, Groves K (eds) Food microstructures: microscopy, measurement and modelling. Woodhead Publishing series in food science technology and nutrition, vol 254. Elsevier Science, Burlington, pp 246–272. <https://doi.org/10.1533/9780857098894.1.246>
- Becker J, Flueckiger R, Reum M, Buechi FN, Marone F, Stampanoni M (2009) Determination of material properties of gas diffusion layers: experiments and simulations using phase contrast tomographic microscopy. *J Electrochem Soc* 156:B1175–B1181. <https://doi.org/10.1149/1.3176876>
- Berujon S, Wang H, Sawhney K (2012) X-ray multimodal imaging using a random-phase object. *Phys Rev A* 86:063813. <https://doi.org/10.1103/PhysRevA.86.063813>
- Berujon S, Wang H, Alcock S, Sawhney K (2014) At-wavelength metrology of hard x-ray mirror using near field speckle. *Opt Express* 22:6438–6446. <https://doi.org/10.1364/OE.22.006438>
- Bie BX, Huang JY, Fan D, Sun T, Fezzaa K, Xiao XH, Qi ML, Luo SN (2017) Orientation-dependent tensile deformation and damage of a T700 carbon fiber/epoxy composite: a synchrotron-based study. *Carbon* 121:127–133. <https://doi.org/10.1016/j.carbon.2017.05.083>
- Blankenburg C, Rack A, Daul C, Ohser J (2017) Torsion estimation of particle paths through porous media observed by in-situ time-resolved microtomography. *J Microsc* 266:141–152. <https://doi.org/10.1111/jmi.12524>
- Bonse U, Hart M (1965) An x-ray interferometer. *Appl Phys Lett* 6:155–156. <https://doi.org/10.1063/1.1754212>
- Born M, Wolf E (1980) Principles of optics: electromagnetic theory of propagation, interference and diffraction of light. Elsevier, Amsterdam
- Bronnikov AV (1999) Reconstruction formulas in phase-contrast tomography. *Opt Commun* 171:239–244. [https://doi.org/10.1016/S0030-4018\(99\)00575-1](https://doi.org/10.1016/S0030-4018(99)00575-1)
- Buffiere JY, Savelli S, Maire E (2000) Characterisation of MMCp and cast aluminium alloys. X-ray tomography in material science. Hermes Science Publications, Paris
- Burvall A, Lundstrom U, Takman PAC, Larsson DH, Hertz HM (2011) Phase retrieval in x-ray phase-contrast imaging suitable for tomography. *Opt Express* 19:10359–10376. <https://doi.org/10.1364/oe.19.010359>

- Carrel M, Beltran MA, Morales VL, Derlon N, Morgenroth E, Kaufmann R, Holzner M (2017) Biofilm imaging in porous media by laboratory x-ray tomography: combining a non-destructive contrast agent with propagation-based phase-contrast imaging tools. *PLoS One* 12:e0180374. <https://doi.org/10.1371/journal.pone.0180374>
- Carroll AJ, van Riessen GA, Balaur E, Dolbnya IP, Tran GN, Peele AG (2017) An iterative method for near-field Fresnel region polychromatic phase contrast imaging. *J Opt* 19. <https://doi.org/10.1088/2040-8986/aa72c4>
- Cedola A, Campi G, Pelliccia D, Bukreeva I, Fratini M, Burghammer M, Rigon L, Arfelli F, Chen RC, Dreossi D, Sodini N, Mohammadi S, Tromba G, Cancedda R, Mastrogiamomo M (2014) Three dimensional visualization of engineered bone and soft tissue by combined x-ray micro-diffraction and phase contrast tomography. *Phys Med Biol* 59:189–201. <https://doi.org/10.1088/0031-9155/59/1/189>
- Chapman D, Thominson W, Johnston RE, Washburn D, Pisano E, Gmur N, Zhong Z, Menk R, Arfelli F, Sayers D (1997) Diffraction enhanced x-ray imaging. *Phys Med Biol* 42:2015–2025. <https://doi.org/10.1088/0031-9155/42/11/001>
- Clauzer JF, Reinsch MW (1992) New theoretical and experimental results in Fresnel optics with applications to matter-wave and x-ray interferometry. *Appl Phys B Lasers Opt* 54:380–395
- Cloetens P, Guigay JP, De Martino C, Baruchel J, Schlenker M (1997) Fractional Talbot imaging of phase gratings with hard x rays. *Opt Lett* 22:1059–1061
- Cloetens P, Ludwig W, Baruchel J, Van Dyck D, Van Landuyt J, Guigay JP, Schlenker M (1999) Holotomography: quantitative phase tomography with micrometer resolution using hard synchrotron radiation x rays. *Appl Phys Lett* 75:2912–2914. <https://doi.org/10.1063/1.125225>
- Coindreau O, Mulat C, Germain C, Lachaud J, Vignoles GL (2011) Benefits of x-ray CMT for the modeling of C/C composites. *Adv Eng Mater* 13:178–185. <https://doi.org/10.1002/adem.201000233>
- Connor DM, Zhong Z (2014) Diffraction-enhanced imaging. *Curr Radiol Rep* 2:55. <https://doi.org/10.1007/s40134-014-0055-y>
- Coolidge W (1917) X-ray tube. US Patent, US1211092A
- David C, Nohammer B, Solak H-H, Ziegler E (2002) Differential x-ray phase contrast imaging using a shearing interferometer. *Appl Phys Lett* 81:3287–3289
- Diemoz PC, Endrizzi M, Zapata CE, Pešić ZD, Rau C, Bravin A, Robinson IK, Olivo A (2013) X-ray phase-contrast imaging with nanoradian angular resolution. *Phys Rev Lett* 110:138105
- Diemoz PC, Vittoria FA, Hagen CK, Endrizzi M, Coan P, Brun E, Wagner UH, Rau C, Robinson IK, Bravin A et al (2015) Single-image phase retrieval using an edge illumination x-ray phase-contrast imaging setup. *J Synchrotron Radiat* 22:1072–1077
- Diemoz PC, Hagen CK, Endrizzi M, Minuti M, Bellazzini R, Urbani L, De Coppi P, Olivo A (2017) Single-shot x-ray phase-contrast computed tomography with nonmicrofocal laboratory sources. *Phys Rev Appl* 7:044029
- Donepudi VR, Cesareo R, Brunetti A, Zhong Z, Yuasa T, Akatsuka T, Takeda T, Gigante GE (2010) Cork embedded internal features and contrast mechanisms with Dei using 18, 20, 30, 36, and 40 keV synchrotron x-rays. *Res Nondestruct Eval* 21:171–183. <https://doi.org/10.1080/09349847.2010.493990>
- Du Y, Liu X, Huang J, Lei Y, Zhao Z, Lin D, Guo J, Li J, Niu H (2015) Sampling grating approach for x-ray differential phase contrast imaging. *Opt Express* 23:12712–12719. <https://doi.org/10.1364/OE.23.012712>
- Duke D, Swantek A, Kastengren A, Fezzaa K, Powell C (2015) Recent developments in x-ray diagnostics for cavitation. *SAE Int J Fuels Lubr* 8:135–146. <https://doi.org/10.4271/2015-01-0918>
- Eastwood DS, Bradley RS, Tariq F, Cooper SJ, Taiwo OO, Gelb J, Merkle A, Brett DJL, Brandon NP, Withers PJ, Lee PD, Shearing PR (2014) The application of phase contrast x-ray techniques for imaging Li-ion battery electrodes. *Nucl Instrum Methods Phys Res Sect B* 324:118–123. <https://doi.org/10.1016/j.nimb.2013.08.066>

- Eberhardt SH, Marone F, Stampanoni M, Buchi FN, Schmidt TJ (2014) Quantifying phosphoric acid in high-temperature polymer electrolyte fuel cell components by x-ray tomographic microscopy. *J Synchrotron Radiat* 21:1319–1326. <https://doi.org/10.1107/s1600577514016348>
- Einarsdottir H, Emerson MJ, Clemmensen LH, Scherer K, Willer K, Bech M, Larsen R, Ersbøll BK, Pfeiffer F (2016) Novelty detection of foreign objects in food using multi-modal x-ray imaging. *Food Control* 67:39–47. <https://doi.org/10.1016/j.foodcont.2016.02.023>
- Endrizzi M (2018) X-ray phase-contrast imaging. *Nucl Instrum Methods Phys Res Sect A* 878:88–98. <https://doi.org/10.1016/j.nima.2017.07.036>
- Endrizzi M, Olivo A (2014) Absorption, refraction and scattering retrieval with an edge-illumination-based imaging setup. *J Phys D Appl Phys* 47:505102
- Endrizzi M, Diemoz PC, Millard TP, Louise Jones J, Speller RD, Robinson IK, Olivo A (2014) Hard x-ray dark-field imaging with incoherent sample illumination. *Appl Phys Lett* 104:024106
- Endrizzi M, Basta D, Olivo A (2015a) Laboratory-based x-ray phase-contrast imaging with misaligned optical elements. *Appl Phys Lett* 107:124103
- Endrizzi M, Murat BIS, Fromme P, Olivo A (2015b) Edge-illumination x-ray dark-field imaging for visualising defects in composite structures. *Compos Struct* 134:895–899
- Endrizzi M, Vittoria F, Olivo A (2018) Single-shot x-ray phase retrieval through hierarchical data analysis and a multi-aperture analyser. *J Imaging* 4:76
- Fatima A, Kulkarni VK, Banda NR, Agrawal AK, Singh B, Sarkar PS, Tripathi S, Shripathi T, Kashyap Y, Sinha A (2016) Non-destructive evaluation of teeth restored with different composite resins using synchrotron based micro-imaging. *J Xray Sci Technol* 24:119–132. <https://doi.org/10.3233/xst-160530>
- Fioravanti M, Di Giulio G, Signorini G, Rognoni GR, Sodini N, Tromba G, Zanini F (2017) Non-invasive wood identification of historical musical bows. *IAWA J* 38:285–296. <https://doi.org/10.1163/22941932-20170172>
- Franco M, Yokaichiya F, Kardjilov N, Ferraz ACD (2015) Microfocus x-ray imaging of Brazil nuts for quality control. *Semina Cienc Agrar* 36:2565–2575. <https://doi.org/10.5433/1679-0359.2015v36n4p2565>
- Fu J, Liu C (2016) In-line phase contrast computed tomography of carbon/carbon composites. In: 2016 IEEE international conference on information and automation
- Gabor D (1948) A new microscopic principle. *Nature* 161:777–778. <https://doi.org/10.1038/161777a0>
- Garcia-Moreno F, Kamm PH, Neu T, Heim K, Rack A, Banhart J (2017) In situ x-ray tomography of aqueous foams: analysis of columnar foam generation. *Colloids Surf A Physicochem Eng Asp* 534:78–84. <https://doi.org/10.1016/j.colsurfa.2017.03.011>
- Gkoumas S, Wang ZT, Abis M, Arboleda C, Tudosie G, Donath T, Bronnimann C, Schulze-Briess C, Stampanoni M (2016) Grating-based interferometry and hybrid photon counting detectors: towards a new era in x-ray medical imaging. *Nucl Instrum Methods Phys Res Sect A* 809:23–30. <https://doi.org/10.1016/j.nima.2015.08.017>
- Goetz K, Kalashnikov MP, Mikhailov YA, Sklizkov GV, Fedotov SI, Foerster E, Zaumseil P (1979) Measurements of the parameters of shell targets for laser thermonuclear fusion using an x-ray Schlieren method. *Sov J Quantum Electron* 9:607
- Greenfeld I, Fezzaa K, Rafailovich MH, Zussman E (2012) Fast x-ray phase-contrast imaging of electrospinning polymer jets: measurements of radius, velocity, and concentration. *Macromolecules* 45:3616–3626. <https://doi.org/10.1021/ma300237j>
- Gresil M, Revol V, Kitsianos K, Kanderakis G, Koulalis I, Sauer MO, Tretout H, Madrigal AM (2017) EVITA project: comparison between traditional non-destructive techniques and phase contrast x-ray imaging applied to aerospace carbon fibre reinforced polymer. *Appl Compos Mater* 24:513–524. <https://doi.org/10.1007/s10443-016-9540-1>
- Gui J-Y, Zhou B, Zhong Y-H, Du A, Shen J (2011) Fabrication of gradient density SiO(2) aerogel. *J Sol-Gel Sci Technol* 58:470–475. <https://doi.org/10.1007/s10971-011-2415-x>
- Guigay JP (1977) Fourier-transform analysis of Fresnel diffraction patterns and in-line holograms. *Optik* 49:121–125

- Guo EY, Zeng G, Kazantsev D, Rockett P, Bent J, Kirkland M, Van Dalen G, Eastwood DS, St. John D, Lee PD (2017) Synchrotron x-ray tomographic quantification of microstructural evolution in ice cream – a multiphase soft solid. *RSC Adv* 7:15561–15573. <https://doi.org/10.1039/c7ra00642j>
- Gureyev TE, Evans R, Stevenson AW, Wilkins SW, Appita (1999) X-ray phase-contrast microscopy of wood and paper. In: 53rd Appita annual conference, proceedings, vols 1 and 2
- Hagen CK, Diemoz PC, Endrizzi M, Rigon L, Dreossi D, Arfelli F, Lopez FCM, Longo R, Olivo A (2014a) Theory and preliminary experimental verification of quantitative edge illumination x-ray phase contrast tomography. *Opt Express* 22:7989–8000
- Hagen CK, Munro PRT, Endrizzi M, Diemoz PC, Olivo A (2014b) Low-dose phase contrast tomography with conventional x-ray sources. *Med Phys* 41:070701
- Hagen CK, Maghsoudlou P, Totonelli G, Diemoz PC, Endrizzi M, Rigon L, Menk R-H, Arfelli F, Dreossi D, Brun E et al (2015) High contrast microstructural visualization of natural acellular matrices by means of phase-based x-ray tomography. *Sci Rep* 5:18156
- Heycock CT, Neville FH (1898) Rontgen ray photography and alloys. *J Chem Soc* 73:714–723
- Hofmann R, Moosmann J, Baumbach T (2011) Criticality in single-distance phase retrieval. *Opt Express* 19:25881–25890. <https://doi.org/10.1364/oe.19.025881>
- Holmstad R, Goel A, Ramaswamy S, Gregersen OW (2006) Visualization and characterization of high resolution 3D images of paper samples. *Appita J* 59:370–377
- Hu ZW, De Carlo F (2008) Noninvasive three-dimensional visualization of defects and crack propagation in layered foam structures by phase-contrast microimaging. *Scr Mater* 59:1127–1130. <https://doi.org/10.1016/j.scriptamat.2008.07.043>
- Hu ZH, Sun M, Lv M, Wang LH, Shi JY, Xiao TQ, Cao Y, Wang J, Fan CH (2016) Deciphering buried air phases on natural and bioinspired superhydrophobic surfaces using synchrotron radiation-based x-ray phase-contrast imaging. *NPG Asia Mater* 8. <https://doi.org/10.1038/am.2016.122>
- Hudspeth M, Claus B, Dubelman S, Black J, Mondal A, Parab N, Funnell C, Hai F, Qi ML, Fezzaa K, Luo SN, Chen W (2013) High speed synchrotron x-ray phase contrast imaging of dynamic material response to split Hopkinson bar loading. *Rev Sci Instrum* 84:025102. <https://doi.org/10.1063/1.4789780>
- Izadifar Z, Chapman LD, Chen XB (2014) Computed tomography diffraction-enhanced imaging for in situ visualization of tissue scaffolds implanted in cartilage. *Tissue Eng Part C Methods* 20:140–148. <https://doi.org/10.1089/ten.tec.2013.0138>
- Jacobsen C, Howells M, Kirz J, Rothman S (1990) X-ray holographic microscopy using photoresists. *J Opt Soc Am A Opt Image Sci Vis* 7:1847–1861. <https://doi.org/10.1364/JOSAA.7.001847>
- Jerjen I, Revol V, Brunner AJ, Schuetz P, Kottler C, Kaufmann R, Lue thi T, Nicoletti G, Urban C, Sennhauser U (2013) Detection of stress whitening in plastics with the help of x-ray dark field imaging. *Polym Test* 32:1094–1098. <https://doi.org/10.1016/j.polymertesting.2013.06.008>
- Johanson Z, Boisvert C, Maksimenko A, Currie P, Trinajstic K (2015) Development of the Synarcual in the Elephant Sharks (Holocephali; Chondrichthyes): implications for vertebral formation and fusion. *PLoS One* 10. <https://doi.org/10.1371/journal.pone.0135138>
- Kagias M, Wang Z, Villanueva-Perez P, Jefimovs K, Stampanoni M (2016) 2D-omnidirectional hard-x-ray scattering sensitivity in a single shot. *Phys Rev Lett* 116:093902. <https://doi.org/10.1103/PhysRevLett.116.093902>
- Kallion GK, Wesolowski M, Vittoria FA, Endrizzi M, Basta D, Millard TP, Diemoz PC, Olivo A (2015) A laboratory based edge-illumination x-ray phase-contrast imaging setup with two-directional sensitivity. *Appl Phys Lett* 107:204105
- Kashyap YS, Agrawal A, Sarkar PS, Shukla M, Roy T, Sinha A (2011) Study of pyro-carbon coated alumina kernel using mixed contrast transfer based x-ray phase retrieval technique. *NDT&E Int* 44:41–46. <https://doi.org/10.1016/j.ndteint.2010.09.004>
- Kastner J, Plank B, Requena G (2012) Non-destructive characterisation of polymers and Al-alloys by polychromatic cone-beam phase contrast tomography. *Mater Charact* 64:79–87. <https://doi.org/10.1016/j.matchar.2011.12.004>

- Khlifa I, Vabre A, Hocevar M, Fezzaa K, Fuzier S, Roussette O, Coutier-Delgosha O (2017) Fast x-ray imaging of cavitating flows. *Exp Fluids* 58. <https://doi.org/10.1007/s00348-017-2426-7>
- Kobayashi T, Toda H (2007) Strength and fracture of aluminium alloys. In: Chandra TTKMMRC (ed) THERMEC 2006, Pts 1–5, vols 539–543. Materials Science Forum, pp 127–134
- Kohn VG, Argunova TS, Je JH (2014) Capsule-like voids in SiC single crystal: phase contrast imaging and computer simulations. *AIP Adv* 4. <https://doi.org/10.1063/1.4896512>
- Kono Y, Kenney-Benson C, Shibasaki Y, Park C, Wang YB, Shen GY (2015) X-ray imaging for studying behavior of liquids at high pressures and high temperatures using Paris-Edinburgh press. *Rev Sci Instrum* 86. <https://doi.org/10.1063/1.4927227>
- Kostenko A, Sharma H, Dere EG, King A, Ludwig W, van Oel W, Stallinga S, van Vliet LJ, Offerman SE (2012) Three-dimensional morphology of cementite in steel studied by x-ray phase-contrast tomography. *Scr Mater* 67:261–264. <https://doi.org/10.1016/j.scriptamat.2012.04.034>
- Kozioziemski BJ, Koch JA, Barty A, Martz HE, Lee WK, Fezzaa K (2005) Quantitative characterization of inertial confinement fusion capsules using phase contrast enhanced x-ray imaging. *J Appl Phys* 97:063103. <https://doi.org/10.1063/1.1862764>
- Lee JS, Park SJ, Lee JH, Weon BM, Fezzaa K, Je JH (2015) Origin and dynamics of vortex rings in drop splashing. *Nat Commun* 6. <https://doi.org/10.1038/ncomms9187>
- Li T, Fan D, Lu L, Huang JY, Zhao F, Qi ML, Sun T, Fezzaa K, Xiao XH, Zhou XM, Suo T, Chen W, Li YL, Zhu MH, Luo SN (2015) Dynamic fracture of C/SiC composites under high strain-rate loading: microstructures and mechanisms. *Carbon* 91:468–478. <https://doi.org/10.1016/j.carbon.2015.05.015>
- Li HY, Kingston AM, Myers GR, Beeching L, Sheppard AP (2018) Linear iterative near-field phase retrieval (LIPR) for dual-energy x-ray imaging and material discrimination. *J Opt Soc Am A Opt Image Sci Vis* 35:A30–A39. <https://doi.org/10.1364/josaa.35.000a30>
- Lohmann M, Dix W, Metge J, Reime B, Schlüter J, Vogel B, Vogel H (2002) HASYLAB annual report 2002. [http://hasyweb.desy.de/science/annual\\_reports/2002\\_report/](http://hasyweb.desy.de/science/annual_reports/2002_report/)
- Lomas H, Jenkins DR, Mahoney MR, Pearce R, Roest R, Steel K, Mayo S (2017) Examining mechanisms of metallurgical coke fracture using micro-CT imaging and analysis. *Fuel Process Technol* 155:183–190. <https://doi.org/10.1016/j.fuproc.2016.05.039>
- Maksimcuka J, Obata A, Sampson WW, Blanc R, Gao CX, Withers PJ, Tsigkou O, Kasuga T, Lee PD, Poologasundarampillai G (2017) X-ray tomographic imaging of tensile deformation modes of electrospun biodegradable polyester fibers. *Front Mater* 4. <https://doi.org/10.3389/fmats.2017.00043>
- Malecki A, Potdevin G, Biernath T, Eggel E, Garcia EG, Baum T, Noel PB, Bauer JS, Pfeiffer F (2013) Coherent superposition in grating-based directional dark-field imaging. *PLoS One* 8. <https://doi.org/10.1371/journal.pone.0061268>
- Malecki A, Eggel E, Schaff F, Potdevin G, Baum T, Garcia EG, Bauer JS, Pfeiffer F (2014) Correlation of x-ray dark-field radiography to mechanical sample properties. *Microsc Microanal* 20:1528–1533. <https://doi.org/10.1017/s1431927614001718>
- Marenzana M, Hagen CK, Borges PDN, Endrizzi M, Szafraniec MB, Vincent TL, Rigon L, Arfelli F, Menk R-H, Olivo A (2014) Synchrotron-and laboratory-based x-ray phase-contrast imaging for imaging mouse articular cartilage in the absence of radiopaque contrast agents. *Philos Trans R Soc Lond A* 372:20130127
- Mason-Smith N, Duke DJ, Kastengren AL, Traini D, Young PM, Chen Y, Lewis DA, Edgington-Mitchell D, Honnery D (2017) Revealing pMDI spray initial conditions: flashing, atomisation and the effect of ethanol. *Pharm Res* 34:718–729. <https://doi.org/10.1007/s11095-017-2098-2>
- Mayo SC, Stevenson AW, Wilkins SW (2012) In-line phase-contrast x-ray imaging and tomography for materials science. *Materials* 5:937–965. <https://doi.org/10.3390/ma5050937>
- Mayo SC, McCann T, Day L, Favaro J, Tuhumury H, Thompson D, Maksimenko A (2016) Rising dough and baking bread at the Australian synchrotron. In: DeJonge MD, Paterson DJ, Ryan CG (eds) Xrm 2014: proceedings of the 12th international conference on x-ray microscopy. AIP conference proceedings, vol 1696. <https://doi.org/10.1063/1.4937500>
- Messe O, Lachambre J, King A, Buffiere JY, Rae CMF (2014) Investigation of fatigue crack propagation in nickel superalloy using diffraction contrast tomography and phase contrast

- tomography. In: Clark G, Wang CH (eds) 11th international fatigue congress, Pts 1 and 2, vols 891–892. Advanced Materials Research, pp 923–928. <https://doi.org/10.4028/www.scientific.net/AMR.891-892.923>
- Miklos R, Nielsen MS, Einarsdottir H, Lametsch R (2016) Grating-based x-ray tomography of 3D food structures. In: Chinesta F, Cueto E, Abisset-Chavanne E (eds) Proceedings of the 19th international ESAFORM conference on material forming. AIP conference proceedings, vol 1769. <https://doi.org/10.1063/1.4963604>
- Miller EA, White TA, McDonald BS, Seifert A (2013) Phase contrast x-ray imaging signatures for security applications. *IEEE Trans Nucl Sci* 60:416–422. <https://doi.org/10.1109/tns.2012.2227803>
- Miyagi M, Kawahito Y, Kawakami H, Shoubu T (2017) Dynamics of solid-liquid interface and porosity formation determined through x-ray phase-contrast in laser welding of pure Al. *J Mater Process Technol* 250:9–15. <https://doi.org/10.1016/j.jmatprotec.2017.06.033>
- Mocella V, Brun E, Ferrero C, Delattre D (2015) Revealing letters in rolled Herculaneum papyri by x-ray phase-contrast imaging. *Nat Commun* 6. <https://doi.org/10.1038/ncomms6895>
- Modregger P, Cremona TP, Benarafa C, Schittny JC, Olivo A, Endrizzi M (2016) Small angle x-ray scattering with edge-illumination. *Sci Rep* 6:30940
- Modregger P, Kagias M, Irvine SC, Brönnimann R, Jefimovs K, Endrizzi M, Olivo A (2017) Interpretation and utility of the moments of small-angle x-ray scattering distributions. *Phys Rev Lett* 118:265501
- Momose A (2005) Recent advances in x-ray phase imaging. *Jpn J Appl Phys* 44:6355
- Momose A, Kawamoto S, Koyama I, Hamaishi Y, Takai K, Suzuki Y (2003) Demonstration of x-ray Talbot interferometry. *Jpn J Appl Phys* 42:L866
- Moon S, Komada K, Sato K, Yokohata H, Wada Y, Yasuda N (2015) Ultrafast x-ray study of multi-hole GDI injector sprays: effects of nozzle hole length and number on initial spray formation. *Exp Thermal Fluid Sci* 68:68–81. <https://doi.org/10.1016/j.expthermflusci.2015.03.027>
- Moreau JD, Cloetens P, Gomez B, Daviero-Gomez V, Nraudeau D, Lafford TA, Tafforeau P (2014) Multiscale 3D virtual dissections of 100-million-year-old flowers using x-ray synchrotron micro-and nanotomography. *Microsc Microanal* 20:305–312. <https://doi.org/10.1017/s1431927613014025>
- Morgan KS, Paganin DM, Siu KKW (2011) Quantitative single-exposure x-ray phase contrast imaging using a single attenuation grid. *Opt Express* 19:19781–19789. <https://doi.org/10.1364/OE.19.019781>
- Morgan KS, Donnelley M, Farrow N, Fouras A, Yagi N, Suzuki Y, Takeuchi A, Uesugi K, Boucher RC, Siu KKW, Parsons DW (2014) In vivo x-ray imaging reveals improved airway surface hydration after a therapy designed for cystic fibrosis. *Am J Respir Crit Care Med* 190:469–471. <https://doi.org/10.1164/rccm.201405-0855LE>
- Morgan KS, Petersen TC, Donnelley M, Farrow N, Parsons DW, Paganin DM (2016) Capturing and visualizing transient x-ray wavefront topological features by single-grid phase imaging. *Opt Express* 24:24435–24450. <https://doi.org/10.1364/oe.24.024435>
- Munro PRT, Ignatyev K, Speller RD, Olivo A (2012) Phase and absorption retrieval using incoherent x-ray sources. *Proc Natl Acad Sci* 109:13922–13927
- Myers GR, Paganin DM, Gureyev TE, Mayo SC (2008) Phase-contrast tomography of single-material objects from few projections. *Opt Express* 16:908–919. <https://doi.org/10.1364/OE.16.000908>
- Nesterets YI, Gureyev TE, Dimmock MR (2018) Optimisation of a propagation-based x-ray phase-contrast micro-CT system. *J Phys D Appl Phys* 51. <https://doi.org/10.1088/1361-6463/aaacee>
- Nugent KA, Gureyev TE, Cookson DF, Paganin D, Barnea Z (1996) Quantitative phase imaging using hard x rays. *Phys Rev Lett* 77:2961–2964. <https://doi.org/10.1103/PhysRevLett.77.2961>
- Olivo A, Speller R (2006) Experimental validation of a simple model capable of predicting the phase contrast imaging capabilities of any x-ray imaging system. *Phys Med Biol* 51:3015–3030. <https://doi.org/10.1088/0031-9155/51/12/001>
- Olivo A, Speller R (2007) A coded-aperture technique allowing x-ray phase contrast imaging with conventional sources. *Appl Phys Lett* 91:074106

- Olivo A, Arfelli F, Cantatore G, Longo R, Menk RH, Pani S, Prest M, Poropat P, Rigon L, Tromba G et al (2001) An innovative digital imaging set-up allowing a low-dose approach to phase contrast applications in the medical field. *Med Phys* 28:1610–1619
- Olivo A, Bohndiek S, Griffiths J, Konstantinidis A, Speller R (2009) A non-free-space propagation x-ray phase contrast imaging method sensitive to phase effects in two directions simultaneously. *Appl Phys Lett* 94:044108
- Olivo A, Gkoumas S, Endrizzi M, Hagen CK, Szafraniec MB, Diemoz PC, Munro PRT, Ignatyev K, Johnson B, Horrocks JA et al (2013) Low-dose phase contrast mammography with conventional x-ray sources. *Med Phys* 40:090701
- Olubamiji AD, Izadifar Z, Chen DX (2014) Synchrotron imaging techniques for bone and cartilage tissue engineering: potential, current trends, and future directions. *Tissue Eng Part B Rev* 20:503–522. <https://doi.org/10.1089/ten.teb.2013.0493>
- Paganin D, Mayo SC, Gureyev TE, Miller PR, Wilkins SW (2002) Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object. *J Microsc (Oxford)* 206:33–40. <https://doi.org/10.1046/j.1365-2818.2002.01010.x>
- Pamukcu AS, Gualda GAR, Rivers ML (2013) Quantitative 3D petrography using x-ray tomography 4: assessing glass inclusion textures with propagation phase-contrast tomography. *Geosphere* 9:1704–1713. <https://doi.org/10.1130/ges00915.1>
- Parab ND, Black JT, Claus B, Hudspeth M, Sun JZ, Fezzaa K, Chen WNW (2014) Observation of crack propagation in glass using x-ray phase contrast imaging. *Int J Appl Glass Sci* 5:363–373. <https://doi.org/10.1111/ijag.12092>
- Paris JL, Kamke FA, Xiao XH (2015) X-ray computed tomography of wood-adhesive bondlines: attenuation and phase-contrast effects. *Wood Sci Technol* 49:1185–1208. <https://doi.org/10.1007/s00226-015-0750-8>
- Parish RW (1986) Microfocus x-ray technology- a review of developments and application. In: Thompson DO, Chimenti DE (eds) Review of progress in quantitative nondestructive evaluation, Vol 5A, Plenum Press, New York, pp 1–20
- Peris D, Hava J (2016) New species from Late Cretaceous New Jersey amber and stasis in subfamily Attageninae (Insecta: Coleoptera: Dermestidae). *J Paleontol* 90:491–498. <https://doi.org/10.1017/jpa.2016.51>
- Perreau M (2012) Description of a new genus and two new species of Leiodidae (Coleoptera) from Baltic amber using phase contrast synchrotron x-ray microtomography. *Zootaxa* 3455:81–88
- Pfeiffer F (2018) X-ray ptychography. *Nat Photonics* 12:9–17. <https://doi.org/10.1038/s41566-017-0072-5>
- Pfeiffer F, Weitkamp T, Bunk O, David C (2006) Phase retrieval and differential phase-contrast imaging with low-brilliance x-ray sources. *Nat Phys* 2:258–261
- Pfeiffer F, Bech M, Bunk O, Kraft P, Eikenberry EF, Brönnimann Ch, Grünzweig C, David C (2008) Hard-x-ray dark-field imaging using a grating interferometer. *Nat Mater* 7:134–137
- Pietsch P, Wood V (2017) X-ray tomography for lithium ion battery research: a practical guide. In: Clarke DR (ed) Annual review of materials research, vol 47, pp 451–479. <https://doi.org/10.1146/annurev-matsci-070616-123957>
- Pitts KF, McCann TH, Mayo S, Favaro J, Day L (2016) Effect of the sugar replacement by citrus fibre on the physical and structural properties of wheat-corn based extrudates. *Food Bioprocess Technol* 9:1803–1811. <https://doi.org/10.1007/s11947-016-1764-4>
- Prade F, Chabior M, Malm F, Grosse CU, Pfeiffer F (2015) Observing the setting and hardening of cementitious materials by x-ray dark-field radiography. *Cem Concr Res* 74:19–25. <https://doi.org/10.1016/j.cemconres.2015.04.003>
- Rafsanjani A, Stiefel M, Jefimovs K, Mokso R, Derome D, Carmeliet J (2014) Hygroscopic swelling and shrinkage of latewood cell wall micropillars reveal ultrastructural anisotropy. *J R Soc Interface* 11. <https://doi.org/10.1098/rsif.2014.0126>
- Rao DV, Bhaskaraiah M, Cesareo R, Brunetti A, Akatsuka T, Yuasa T, Zhong Z, Takeda T, Gigante GE (2013) Synchrotron-based non-destructive diffraction-enhanced imaging systems to image walnut at 20 keV. *J Food Meas Charact* 7:13–21. <https://doi.org/10.1007/s11694-012-9134-z>

- Reischig P, Helfen L, Wallert A, Baumbach T, Dik J (2013) High-resolution non-invasive 3D imaging of paint microstructure by synchrotron-based x-ray laminography. *Appl Phys A Mater Sci Process* 111:983–995. <https://doi.org/10.1007/s00339-013-7687-2>
- Revol V, Plank B, Kaufmann R, Kastner J, Kottler C, Neels A (2013) Laminate fibre structure characterisation of carbon fibre-reinforced polymers by x-ray scatter dark field imaging with a grating interferometer. *NDT&E Int* 58:64–71. <https://doi.org/10.1016/j.ndteint.2013.04.012>
- Reza S, Pelzer G, Weber T, Frojdli C, Bayer F, Anton G, Rieger J, Thim J, Michel T, Norlin B (2014) Investigation on the directional dark-field signals from paperboards using a grating interferometer. *J Instrum* 9. <https://doi.org/10.1088/1748-0221/9/04/c04032>
- Roentgen W (1896) On a new kinds of rays. *Science* 3:227–231. <https://doi.org/10.1126/science.3.59.227>
- Rusu LC, Seche E, Freimann PC, Hoinoiu B, Negruiti ML, Ardelean L, Sinescu C (2014) Synchrotron radiation x-ray micro-CT evaluation of bone augmentation. *Rev Chim* 65:1114–1116
- Rutishauser S, Rack A, Weitkamp T, Kayser Y, David C, Macrander AT (2013) Heat bump on a monochromator crystal measured with x-ray grating interferometry. *J Synchrotron Radiat* 20:300–305. <https://doi.org/10.1107/s0909049513001817>
- Sarapata A, Ruiz-Yaniz M, Zanette I, Rack A, Pfeiffer F, Herzen J (2015) Multi-contrast 3D x-ray imaging of porous and composite materials. *Appl Phys Lett* 106. <https://doi.org/10.1063/1.4918617>
- Sasov A, Ceulemans T, van Dyck D (2001) Desk-top x-ray microtomography. In: Tobin KW, Lakhani F (eds) *Metrology-based control for micro-manufacturing*, vol 4275. SPIE – International Society Optical Engineering, Bellingham, pp 147–154
- Schmahl G, Rudolph D, Schneider G, Guttmann P, Niemann B (1994) Phase-contrast x-ray microscopy studies. *Optik* 97:181–182
- Schropp A, Hoppe R, Meier V, Patommel J, Seiboth F, Ping Y, Hicks DG, Beckwith MA, Collins GW, Higginbotham A, Wark JS, Lee HJ, Nagler B, Galtier EC, Arnold B, Zastra U, Hastings JB, Schroer CG (2015) Imaging shock waves in diamond with both high temporal and spatial resolution at an XFEL. *Sci Rep* 5. <https://doi.org/10.1038/srep11089>
- Sinnett-Jones PE, Browne M, Ludwig W, Buffiere JY, Sinclair I (2005) Microtomography assessment of failure in acrylic bone cement. *Biomaterials* 26:6460–6466. <https://doi.org/10.1016/j.biomaterials.2005.04.064>
- Snigirev A, Snigireva I, Kohn V, Kuznetsov S, Schelokov I (1995) On the possibilities of x-ray phase contrast microimaging by coherent high-energy synchrotron radiation. *Rev Sci Instrum* 66:5486–5492. <https://doi.org/10.1063/1.1146073>
- Soriano C, Archer M, Azar D, Creaser P, Delclos X, Godthelp H, Hand S, Jones A, Nel A, Neraudeau D, Ortega-Blanco J, Perez-de la Fuente R, Perrichot V, Saupe E, Kraemer MS, Tafforeau P (2010) Synchrotron x-ray imaging of inclusions in amber. *C R Palevol* 9:361–368. <https://doi.org/10.1016/j.crpv.2010.07.014>
- Stock SR (2008) Recent advances in x-ray microtomography applied to materials. *Int Mater Rev* 53:129–181. <https://doi.org/10.1179/174328008x277803>
- Sun F, Moroni R, Dong K, Markotter H, Zhou D, Hilger A, Zielke L, Zengerle R, Thiele S, Banhart J, Manke I (2017) Study of the mechanisms of internal short circuit in a Li/Li cell by synchrotron x-ray phase contrast tomography. *ACS Energy Lett* 2:94–104. <https://doi.org/10.1021/acsenergylett.6b00589>
- Suortti P, Keyrilainen J, William T (2013) Analyser-based x-ray imaging for biomedical research. *J Phys D Appl Phys* 46:494002
- Tafforeau P, Boistel R, Boller E, Bravin A, Brunet M, Chaimanee Y, Cloetens P, Feist M, Hoszowska J, Jaeger JJ, Kay RF, Lazzari V, Marivaux L, Nel A, Nemoz C, Thibault X, Vignaud P, Zabler S (2006) Applications of x-ray synchrotron microtomography for non-destructive 3D studies of paleontological specimens. *Appl Phys A Mater Sci Process* 83:195–202. <https://doi.org/10.1007/s00339-006-3507-2>
- Takashima K, Hoshino M, Uesugi K, Yagi N, Matsuda S, Nakahira A, Osumi N, Kohzuki M, Onodera H (2015) X-ray phase-contrast computed tomography visualizes the microstructure

- and degradation profile of implanted biodegradable scaffolds after spinal cord injury. *J Synchrotron Radiat* 22:136–142. <https://doi.org/10.1107/s160057751402270x>
- Takeya S, Yoneyama A, Ueda K, Mimachi H, Takahashi M, Sano K, Hyodo K, Takeda T, Gotoh Y (2012) Anomalously preserved clathrate hydrate of natural gas in pellet form at 253 K. *J Phys Chem C* 116:13842–13848. <https://doi.org/10.1021/0302269v>
- Takeya S, Yoneyama A, Ueda K, Hyodo K, Yamawaki H, Fujihisa H, Gotoh Y, Takeda T (2013) Phase-contrast x-ray images of ice and water on carbon paper for fuel cells measured by diffraction-enhanced imaging technique. *Jpn J Appl Phys* 52. <https://doi.org/10.7567/jjap.52.048002>
- Talbot HF (1836) LXXVI. Facts relating to optical science. No. IV. *Lond Edinb Philos Mag J Sci* 9:401–407
- Teague MR (1983) Deterministic phase-retrieval – a Green-function solution. *J Opt Soc Am* 73:1434–1441. <https://doi.org/10.1364/josa.73.001434>
- Toda H, Tomizato F, Harasaki R, Seo D, Kobayashi M, Takeuchi A, Uesugi K (2016) 3D fracture behaviours in dual-phase stainless steel. *ISIJ Int* 56:883–892. <https://doi.org/10.2355/isijinternational.ISIJINT-2015-631>
- Trtik P, Dual J, Keunecke D, Mannes D, Niemz P, Stahli P, Kaestner A, Groso A, Stampanoni M (2007) 3D imaging of microstructure of spruce wood. *J Struct Biol* 159:46–55. <https://doi.org/10.1016/j.jsb.2007.02.003>
- Turner LD, Dhal BB, Hayes JP, Mancuso AP, Nugent KA, Paterson D, Scholten RE, Tran CQ, Peele AG (2004) X-ray phase imaging: demonstration of extended conditions with homogeneous objects. *Opt Express* 12:2960–2965. <https://doi.org/10.1364/opex.12.002960>
- Uehara M, Yashiro W, Momose A (2013) Effectiveness of x-ray grating interferometry for non-destructive inspection of packaged devices. *J Appl Phys* 114. <https://doi.org/10.1063/1.4823982>
- Vavrik D, Jakubek J, Jandejsek I, Krejci F, Kumpova I, Zemlicka J (2015) Visualization of delamination in composite materials utilizing advanced x-ray imaging techniques. *J Instrum* 10. <https://doi.org/10.1088/1748-0221/10/04/c04012>
- Vittoria FA, Kallon GKN, Basta D, Diemoz PC, Robinson IK, Olivo A, Endrizzi M (2015) Beam tracking approach for single-shot retrieval of absorption, refraction, and dark-field signals with laboratory x-ray sources. *Appl Phys Lett* 106. <https://doi.org/10.1063/1.4922189>
- Vittoria FA, Endrizzi M, Kallon GK, Hagen CK, Iacoviello F, De Coppi P, Olivo A (2017) Multimodal phase-based x-ray microtomography with nonmicrofocal laboratory sources. *Phys Rev Appl* 8. <https://doi.org/10.1103/PhysRevApplied.8.064009>
- Vladimirov P, Ferrero C, Chakin V, Kurinskiy P, Moeslang A, Pieritz R, Weitkamp T, Brun E (2015) Microstructure of out-of-pile annealed neutron irradiated beryllium studied by x-ray tomography. *Acta Mater* 88:293–301. <https://doi.org/10.1016/j.actamat.2015.01.045>
- Wagner A, Sachse A, Keller M, Aurich M, Wetzel WD, Hortschansky P, Schmuck K, Lohmann M, Reime B, Metge J, Arfelli F, Menk R, Rigon L, Muehleman C, Bravin A, Coan P, Mollenhauer J (2006) Qualitative evaluation of titanium implant integration into bone by diffraction enhanced imaging. *Phys Med Biol* 51:1313–1324. <https://doi.org/10.1088/0031-9155/51/5/019>
- Wang K, Lei HL, Li J, Lin W, Qi XB, Tang YJ, Liu YQ (2014) Characterization of inertial confinement fusion targets using x-ray phase contrast imaging. *Opt Commun* 332:9–13. <https://doi.org/10.1016/j.optcom.2014.06.066>
- Wang SX, Hu RF, Gao K, Wali F, Zan GB, Wang DJ, Pan ZY, Wei SQ (2017) Non-destructive study of fruits using grating-based x-ray imaging. *Nucl Sci Tech* 28. <https://doi.org/10.1007/s41365-016-0169-4>
- Weitkamp T, Diaz A, David C, Pfeiffer F, Stampanoni M, Cloetens P, Ziegler E (2005) X-ray phase imaging with a grating interferometer. *Opt Express* 13:6296–6304
- Wen H, Bennett EE, Hegedus MM, Rapacchi S (2009) Fourier x-ray scattering radiography yields bone structural information. *Radiology* 251:910–918. <https://doi.org/10.1148/radiol.2521081903>

- Wen HH, Bennett EE, Kopace R, Stein AF, Pai V (2010) Single-shot x-ray differential phase-contrast and diffraction imaging using two-dimensional transmission gratings. *Opt Lett* 35:1932–1934. <https://doi.org/10.1364/OL.35.001932>
- Wilkins SW, Gureyev TE, Gao D, Pogany A, Stevenson AW (1996) Phase-contrast imaging using polychromatic hard x-rays. *Nature* 384:335–338. <https://doi.org/10.1038/384335a0>
- Wu Y, Takano H, Momose A (2017) In situ observation of polymer blend phase separation by x-ray Talbot-Lau interferometer. In: Proceedings of the SPIE – Developments in X-ray tomography XI, vol 10391
- Xiao XH, Fusseis F, De Carlo F (2012) X-ray fast tomography and its applications in dynamical phenomena studies in geosciences at advanced photon source. In: Stock SR (ed) Developments in x-ray tomography viii, vol 8506. Proceedings of SPIE. <https://doi.org/10.1117/12.936331>
- Xue Y, Xiao T, Du G, Tong Y, Liu H, Deng B, Xie H, Xu H (2013) Observation of cavitation and water-refilling processes in plants with x-ray phase contrast microscopy. *Nucl Sci Tech* 24:060101
- Yin ZJ, Zhu MY, Bottjer DJ, Zhao FC, Tafforeau P (2016) Meroblastic cleavage identifies some Ediacaran Doushantuo (China) embryo-like fossils as metazoans. *Geology* 44:735–738. <https://doi.org/10.1130/g38262.1>
- Young ML, Rao R, Almer JD, Haefner DR, Lewis JA, Dunand DC (2009) Effect of ceramic preform geometry on load partitioning in Al<sub>2</sub>O<sub>3</sub>-Al composites with three-dimensional periodic architecture. *Mater Sci Eng A* 526:190–196. <https://doi.org/10.1016/j.msea.2009.07.033>
- Zabler S, Rack T, Rack A, Nelson K (2012) Fatigue induced deformation of taper connections in dental titanium implants. *Int J Mater Res* 103:207–216. <https://doi.org/10.3139/146.110666>
- Zamir A, Diemoz PC, Vittoria FA, Hagen CK, Endrizzi M, Olivo A (2017) Edge illumination x-ray phase tomography of multi-material samples using a single-image phase retrieval algorithm. *Opt Express* 25:11984–11996
- Zanette I, Weitkamp T, Donath T, Rutishauser S, David C (2010) Two-dimensional x-ray grating interferometer. *Phys Rev Lett* 105:248102
- Zanette I, Zhou T, Burvall A, Lundstrom U, Larsson DH, Zdora M, Thibault P, Pfeiffer F, Hertz HM (2014) Speckle-based x-ray phase-contrast and dark-field imaging with a laboratory source. *Phys Rev Lett* 112. <https://doi.org/10.1103/PhysRevLett.112.253903>
- Zbib MB, Parab ND, Chen WNW, Bahr DF (2015) New pulverization parameter derived from indentation and dynamic compression of brittle microspheres. *Powder Technol* 283:57–65. <https://doi.org/10.1016/j.powtec.2015.04.066>
- Zdora MC, Thibault P, Deyhle H, Vila-Comamala J, Rau C, Zanette I (2018) Tunable x-ray speckle-based phase-contrast and dark-field imaging using the unified modulated pattern analysis approach. *J Instrum* 13. <https://doi.org/10.1088/1748-0221/13/05/c05005>
- Zoofan B, Kim JY, Rokhlin SI, Frankel GS (2006) Phase-contrast x-ray imaging for nondestructive evaluation of materials. *J Appl Phys* 100:014502. <https://doi.org/10.1063/1.2209889>



# X-Ray Tomography

31

Johann Kastner and Christoph Heinzl

## Contents

Introduction .....	1096
Principles of X-Ray Computed Tomography Technology .....	1098
Generation of X-Rays and Interaction with Matter .....	1098
Principle of Cone Beam XCT .....	1099
Artifacts .....	1104
Artifact Correction Methods .....	1108
Further Important XCT Methods .....	1111
Applications of X-Ray Computed XCT .....	1117
Overview of Applications .....	1117
XCT for Materials Characterization .....	1118
Characterization of Metals .....	1121
XCT for Metrology .....	1123
Quantitative XCT: Advanced XCT Data Processing and Visualization Methods .....	1130
XCT Data Processing, Analysis, and Visualization .....	1130
Examples for Quantitative XCT .....	1134
Quantitative XCT: Combination of XCT and Materials Simulation .....	1143
4DXCT/In Situ XCT .....	1146
Phase-Contrast XCT .....	1150
Propagation-Based Phase-Contrast XCT .....	1151
Talbot-Lau Interferometer Phase-Contrast XCT .....	1155
Trends and Summary .....	1160
References .....	1162

## Abstract

Over the past years, a large number of novel X-ray imaging and data processing methods have been developed. The application areas of X-ray computed tomography (XCT) are highly diverse and extensive, since any material or component

J. Kastner (✉) · C. Heinzl

University of Applied Sciences Upper Austria, Wels, Austria

e-mail: [johann.kastner@fh-ooe.at](mailto:johann.kastner@fh-ooe.at); [christoph.heinzl@fh-wels.at](mailto:christoph.heinzl@fh-wels.at)

may be examined using XCT. The major application areas of XCT in science and industry are found in nondestructive testing, 3D materials characterization, and dimensional measurements (metrology). The nonmedical XCT market is steadily growing, but the full potential of this technique for industrial applications has not been exploited yet. There are many useful XCT applications which still have to be discovered. This chapter provides an overview of the principles of XCT, of drawbacks such as measurement artifacts as well as their correction, of different XCT methods and scanning protocols, as well as of applications of XCT. The focus of this chapter lies on XCT for materials simulation and high-resolution, quantitative, *in situ*, and phase-contrast XCT.

---

## Introduction

The invention of X-ray computed tomography (XCT) may be considered as one of the greatest innovations in the field of medical diagnostics within the last century. The initial use of computed tomography during the 1970s sparked a revolution in the field of medical science due to the enormous benefit of non-invasive diagnostics. Medical X-ray tomography is well represented in literature, e.g., see Hsieh (2003) or Buzug (2008) and still a highly active topic of research.

In contrast to the field of medicine, XCT for nondestructive evaluation and industrial applications was not established as quickly. One of the first XCT users in Europe outside of medical diagnostics was the company Messerschmidt-Boelkow-Blohm (now Airbus Helicopters) in Munich. This company has successfully applied conventional medical XCT scanners for nondestructive testing of industrial components, more specifically helicopter rotor blades, since 1979 (Oster 1999). Up to the 1990s, industrial XCT was either applied for special application cases, when conventional testing methods failed, or it was considered as a complementary method in the field of nondestructive testing. The breakthrough for industrial XCT in Europe happened much later toward the end of the 1990s when the potential of XCT was recognized in automotive industry. Especially the European car industry was an important pioneer, which applied XCT for first part inspection of engine components made of aluminum castings. The following reasons are mainly accountable for the success of XCT for nondestructive evaluation in industry: The rapid development of high-performance computers and software tools for reconstruction, data processing, and visualization of 3D datasets allowed for exploring and analyzing data in previously impossible detail. Furthermore, new flat-panel matrix detectors with a high dynamic range on the basis of amorphous silicon crystals and efficient scintillator materials increasingly replaced X-ray image amplifiers. In particular, the development of micro-focus-XCT with cone beam geometry and comparatively short measurement times increased the acceptance of this technology enormously (Baruchel et al. 2000; Banhart 2008; Hanke et al. 2008; Landis and Keane 2010; Harrer and Kastner 2011; Maire and Withers 2014; Kastner 2016; Kastner et al. 2017a; Carmignato et al. 2018). Table 1 reports on the history of XCT as well as on important milestones in the development of XCT.

**Table 1** Overview of important milestones in the development of X-ray computed tomography

Year	Milestone
1895	Wilhelm Conrad Röntgen discovered a new kind of radiation, which he named the X-rays
1917	Johann Radon demonstrated mathematically that an object could be replicated from an infinite set of its projections
1967	Godfrey Newbold Hounsfield demonstrated the first clinical CT scanner
1979	MBB (now Airbus Helicopters) used medical XCT devices for testing helicopter rotor blades
1983	Start-up of the first industrial XCT scanner with a cobalt-60 radionuclide source developed by BAM
1984	Publication of the <i>Practical Cone Beam Algorithm</i> by Feldkamp, Davis, and Kress
1995	The first XCT systems equipped with matrix detectors were used for geometrical measurements (e.g., aluminum components)
2004	VDI/VDE Society for Metrology and Automation Engineering are drawing up guidelines for geometrical measurement using XCT
2005	Desktop XCT devices with sub-micro-focus tubes and a spatial resolution in the sub-micro range were introduced
2006	Introduction of Talbot-Lau interferometer XCT systems with low brilliance sources that allow measurement of absorption, refraction, and scattering contrast simultaneously
2018	Developments in the fields of <i>in situ</i> XCT, in-line XCT, diffraction contrast XCT, spectral XCT, XCT simulation, XCT data processing, etc.
2018	>30 XCT suppliers including small companies and big international enterprises
2018	>10 XCT software companies for nonmedical XCT applications worldwide
2018	>3000 XCT installations for industrial and scientific applications worldwide
	XCT is widely used for metrology, nondestructive testing, and materials characterization

The current application areas of XCT in the industrial and scientific context are diverse and extensive, since almost any material or component can be examined with XCT. The major application areas of XCT in science and industry are found in nondestructive testing, 3D materials characterization as well as dimensional measurements (metrology). The nonmedical XCT market is a steadily growing market, but the full potential for industry is not yet explored. There are many useful applications of XCT which still have to be discovered. Current estimates for the worldwide industrial XCT market indicate 2000–3000 XCT systems for nonmedical applications in 2017 (Kastner et al. 2017a), which will support the development of novel XCT applications. Nowadays, industrial X-ray computed tomography is regarded as efficient and highly precise method that provides comprehensive 100% inspections and accurate 3D volumetric data of specimens with reasonable size. An important aspect for practical applications of XCT in industry is standardization. Several industrial XCT standards are already available and adopted by industry (Bartscher et al. 2018), while others are still under development:

- VDI/VDE 2630 for metrology (Basic principles, uncertainty/process suitability/calibration, software systems).
- DIN EN 16016-1-4: 2011 for nondestructive testing (Terminology, principle, operation, and qualification).

- ASTM E 1695-95 (Standard test method for measurement of CT system performance) and ASTM E 1441-11 and ASTM E 1570-11 (Standard practice for CT examination), and ASTM E 1672-12 (Standard guide for CT system selection).
- ISO 15708-1 and 2 for nondestructive testing (Principles and examination practices).
- ISO/WD 10360: An international standard defining acceptance and reverification test is currently under development by ISO TC213 WG10. The main purpose of ISO 10360 XCT is to achieve comparability with the characteristics of coordinate measuring systems, and it is dedicated to metrological measurements.

---

## Principles of X-Ray Computed Tomography Technology

### Generation of X-Rays and Interaction with Matter

The conventional method of generating X-rays in a laboratory setting is based on an evacuated X-ray tube. This evacuated tube integrates a cathode, emitting electrons into the vacuum, and an anode collecting the electrons. Using this principle, a flow of electrons is established through the tube, when electric current is applied. Typically, for X-ray sources acceleration voltages of up to 450 kV (a few X-ray tubes operate with voltages up to 800 kV) are used. So, the electrons in the electron cloud formed around the cathode filament are subsequently accelerated toward the anode. The anode is typically made of tungsten, molybdenum, or copper due to their convenient X-ray energy spectrum and due to their high melting points. When the accelerated electrons interact with the target material at the anode, different types of collisions take place and generate an X-ray spectrum with Bremsstrahlung and characteristic radiation. Several fundamental ways in which X-ray photons interact with matter are known, e.g., photoelectric absorption, Rayleigh scattering, Compton scattering, and pair production.

For quantitative description of absorption and scattering of electromagnetic radiation,  $I_0(E)$  is considered as the incident photon flux. These photons interact with the matter of a sample of thickness  $s$  and after absorption and scattering,  $I(E)$  is recorded by a detector. The relationship between  $I_0(E)$  and  $I(E)$  is described in Lambert-Beer's law (Lambert 1760) or exponential attenuation law by Eq. 1.

$$I(E) = I_0(E) \cdot e^{-\int_0^s \mu_{l,s}(E, \rho, Z) \cdot ds} \quad (1)$$

The detection of X-ray photons is based on the same principles of interactions between X-ray radiation and matter as described above. The purpose of a detector is to convert the photon flux  $I(E)$  into an electrical signal. In the first tomographic systems gas detectors, also known as Geiger-Müller counters, were used. For state-of-the-art XCT devices the most common detectors are digital flat-panel detectors or charge-coupled device cameras (Carmignato et al. 2018). X-ray detectors may be divided into direct and indirect conversion detectors. Both types are based on thin

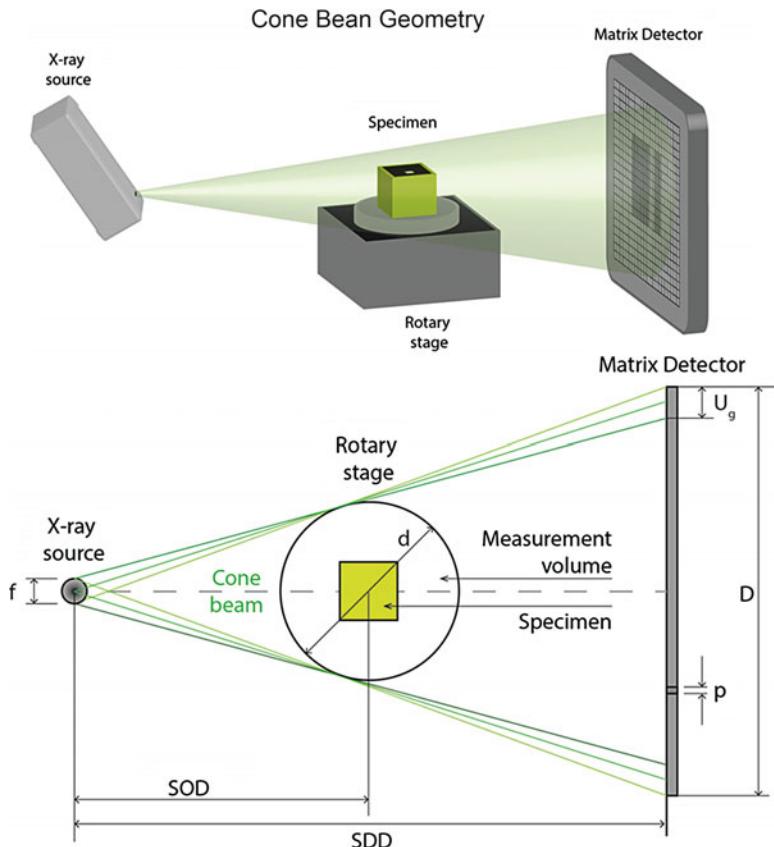
layers of amorphous silicon or other semiconductors deposited onto glass substrates, with two-dimensional arrays of detector elements. In indirect detection systems, X-ray photons collide with a phosphorescent layer (terbium-doped gadolinium oxysulfide ( $\text{Gd}_2\text{O}_2\text{S}:\text{Tb}$ , Gadox) and thallium-doped cesium iodide ( $\text{CsI}:\text{Tl}$ ) are commonly used) and produce visible light, which is in turn detected by an array of thin film transistor-switched photodiodes. In the case of detectors with direct conversion, the radiation is directly converted in the sensor layer which is mostly based on selenium (Se), cadmium telluride (CdTe), or zinc cadmium telluride (ZnCdTe). Direct converting detectors are photon counting sensors, which allow for “single photon counting.” This means that every single photon of X-ray radiation which is detected in a pixel is processed by the detector and counted. This principle features the advantages of high contrast together with sharp images and spectral information of X-rays. Due to the measurement speed and quality, XCT systems with cone beam geometry and indirect converting flat-panel matrix detectors have gained general acceptance for industrial applications and for materials characterization.

## Principle of Cone Beam XCT

The principle of cone beam XCT is shown in Fig. 1. A specimen is placed on a high precision rotation stage between the X-ray tube and the digital matrix detector. The X-ray source generates a conic beam of photons that penetrate the specimen to be analyzed. Due to the cone beam geometry a magnified penetration image (2D radiograph) is recorded by the X-ray matrix detector. Reducing the distance between source and object allows for larger magnifications and an improved spatial resolution, but it also reduces the field of view. The specimen is rotated step by step, taking projection images at each angular position. An XCT scan typically covers step widths of fractions of a degree and a full rotation of  $360^\circ$ .

The achievable resolution of an XCT scan of a specimen depends on the geometrical magnification determined by the position of the object with respect to source and detector. The closer to the source a specimen is positioned, the higher is the magnification and the smaller the voxel size. Such a setup results in a higher-resolution XCT scan. The resolution is also affected by the size of the focal spot (directly proportional to the geometric unsharpness  $U_g$ ), the quality of the rotation axis and other characteristics of the XCT device. For cone beam XCT, the highest accuracy is achieved in the beam which is perpendicular to the rotary axis, going through the focal spot of the X-ray source. The geometrical magnification determines the opening cone beam angle: at high magnifications, the angle increases and also the measurement errors. Geometrical magnification  $m$  is given as a ratio between a source-detector distance (SDD) and a source-object distance (SOD) according to Eq. (2).

$$m = \frac{\text{SDD}}{\text{SOD}} \quad (2)$$



**Fig. 1** Main components of cone beam X-ray computed tomography: X-ray source (X-ray tube), rotary stage with specimen and a matrix detector (flat panel detector) (top image). Typical geometry of cone beam XCT (bottom image, top view): Geometrical magnification is determined by the source-object distance (SOD) and source-detector distance (SDD), as well as the effective detector width  $D$  and the diameter  $d$  of the measuring volume.  $f$  is the focal spot size,  $p$  is the detector pixel size and  $U_g$  the geometric unsharpness in the penetration image in relation to the focal spot size  $f$

The size of a voxel  $s$  is then determined by the detector pixel size  $p$  according to Eq. (3). Typical voxel sizes in the reconstructed XCT are in range between 1  $\mu\text{m}$  and 0.2 mm. For high-resolution XCT systems, the minimum voxel size  $s$  can be 100 nm or even lower.

$$s = \frac{p}{m} \quad (3)$$

The voxel-size  $s$  in the reconstructed XCT data depends on the number of pixels  $N_{\text{pixels}}$  of the X-ray detector and the diameter  $d$  of the scanned sample and may also be computed as follows:

$$s = \frac{N_{\text{pixels}}}{d} \quad (4)$$

In order to accurately reconstruct the volumetric data, the entire sample must remain within the field of view and the cone beam during the rotation. The maximum magnification of an XCT system is limited by the ratio of the detector width D and the sample diameter d. It can be expressed according to Eq. (5).

$$m = \frac{D}{d} \quad (5)$$

The geometrical unsharpness  $U_g$  depends on the focal spot size and on the magnification as presented in Eq. (6). Since the focal spot of the X-ray source has a finite size, geometric unsharpness or image blur will always occur.

$$U_g = f(m - 1) \quad (6)$$

Conventional industrial X-ray tubes feature focal spots of 0.5–4 mm in size, minifocus X-ray tubes 0.1–0.5 mm. Focal spots of industrial micro-focus X-ray tubes range from 1 to 50  $\mu\text{m}$  and nano-focus from 0.4  $\mu\text{m}$  to 1  $\mu\text{m}$ . The total unsharpness is further increased by the detector pixel size, by detector irregularities, geometrical imperfection during the XCT acquisitions procedure and motion unsharpness.

In order to compute a 3D image from the set of 2D penetration images generated in XCT, reconstruction algorithms are used. Feldkamp et al. (1984) describe an approximating reconstruction algorithm to retrieve 3D data from a series of 2D radiographs for circular cone beam tomography (FDK algorithm). The FDK algorithm filters the weighted projection data on each ray and then back projects them into the reconstruction space. Each fan is angled out of the source-detector plane of rotation. In cone beam reconstruction, the fan is tilted out of the plane of rotation; thus the size of the fan and the coordinate system of the reconstructed point are changed. The result is a volumetric dataset, which consists of discrete volumetric elements. A volumetric element is also called “voxel” (cf. pixel = picture element). Each voxel represents a gray value corresponding to the effective X-ray attenuation coefficient at the respective position in the volumetric dataset. The FDK algorithm generates an approximation in the sense that the reconstructed 3D volume will deviate from the ideal as a result of the geometric setup, regardless of the measurement resolution. For moderate cone angles, these differences are however small and often negligible. Yet, the simplicity of FDK turned it into the most commonly used algorithm for cone beam reconstruction.

Besides FDK, there are a large number of other reconstruction techniques such as three-dimensional Fourier-based reconstruction methods, algebraic and statistical reconstruction methods (e.g., algebraic reconstruction technique (ART) and direct iterative reconstruction (DIRECTT)) (Hanke et al. 2008; Carmignato et al. 2018). However, for this chapter reconstruction algorithms are considered as out of scope.

**Table 2** Typical maximum penetrable material thicknesses for common materials

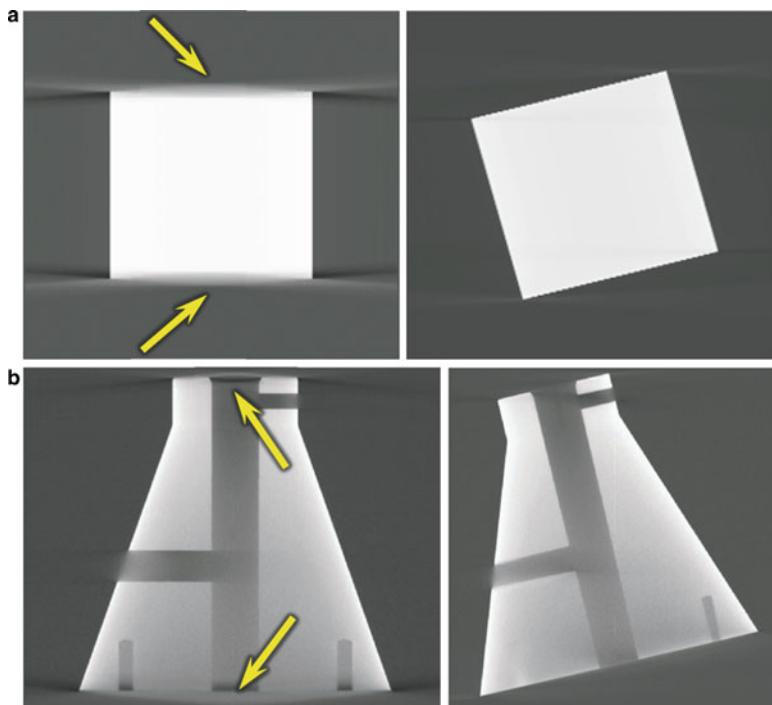
Material vs. X-ray voltage	130 kV	190 kV	225 kV	450 kV
Plastics	90 mm	200 mm	250 mm	450 mm
Aluminum	30 mm	90 mm	150 mm	250 mm
Steel	5 mm	25 mm	40 mm	70 mm

The set of the generated projection data for a given slice through an object is called “sinogram.” A sinogram is presented as an image where the number of lines corresponds to the number of projections taken; each line represents a projection at a given angle and the number of columns is equal to the width of the detector. A sinogram is useful for analyzing the projection data itself and for detecting anomalies in the data acquisition: For example, a defective detector pixel manifests itself as a vertical line in the corresponding sinogram and a temporary malfunction of the X-ray tube produces a disrupted sinogram with horizontal lines (Hsieh 2003).

There are restrictions concerning the maximum accumulated material thickness that can be penetrated by X-rays. The maximum material thickness depends on the material attenuation coefficient and the X-ray photon energy. In Table 2 typical values are given for plastics, aluminum, and steel. The values in this table are maximum accumulated thicknesses producing low signal-to-noise ratios; with decreasing thicknesses, the transmitted intensity increases, and possible measurement artifacts are reduced.

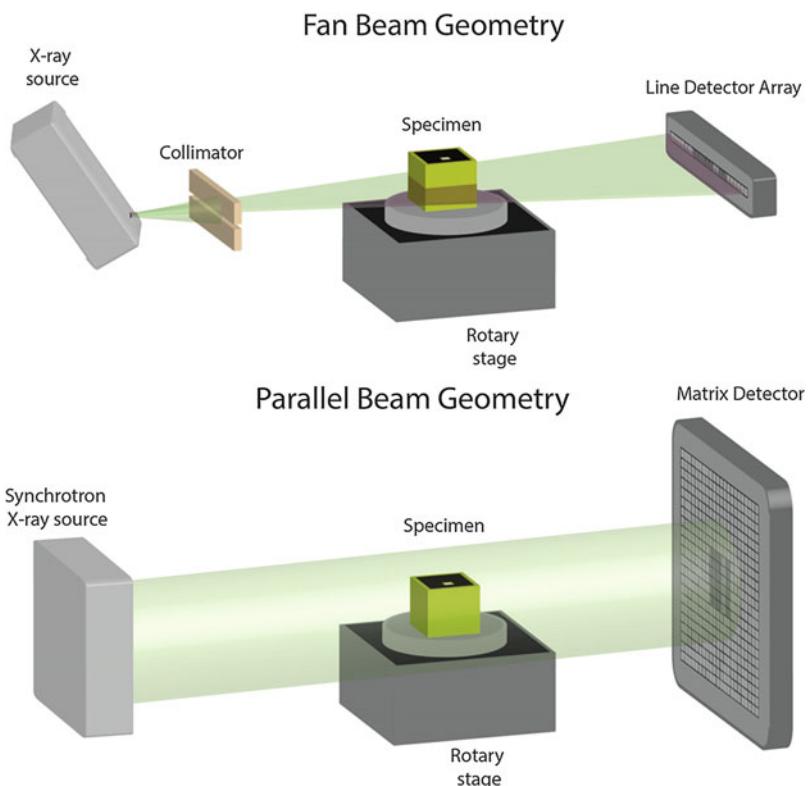
Before scanning, the object should be oriented in a way to minimize the maximum penetrated material thickness in order to reduce artifacts. Due to the incompleteness of the acquisition geometry of cone beam XCT, an exact reconstruction of planar faces parallel to the central beam is not possible without errors (Feldkamp artifacts). For this reason, these faces will appear blurred in the resulting reconstructed volume (see yellow arrows in Fig. 2). Corresponding to Tuy-Smith (1983) the sufficiency condition for full reconstruction is defined as follows: If every plane that intersects the object, at some point during the scanning rotation, contains the X-ray source point, then one can fully reconstruct the object. For an accurate reconstruction of an arbitrary plane, the circular trajectory of the source must be intersected. A face can be accurately reconstructed if its plane has an intersection with the circular trajectory of the source. As we can see from this example, we can get rid of the blurred faces by choosing a proper placement of the specimen; this means a tilted position of about 15° during the XCT scan.

**Fan beam X-ray computed tomography and parallel beam synchrotron X-ray tomography (s-XCT):** Fig. 3 shows further XCT acquisition configurations: fan beam geometry with an X-ray source, a collimator and 1D detector, and parallel beam geometry. When using fan beam XCT, the sample needs vertical translation to scan a 3D volume. The collimator of fan beam XCT systems together with special X-ray line detectors leads to a reduction of scattered radiation and scattering artifacts which is most important for larger metallic parts. Due to the much longer acquisition times, fan beam XCT is not as frequently used as cone beam XCT.



**Fig. 2** Influence of the specimen positioning on reconstructed XCT data. For a cube specimen and a more complex specimen, scans were performed with a vertical positioning (images on the left for specimens **a** and **b**), as well as scan with a tilted position, respectively (images on the right side). Artifacts are marked with yellow arrows. The placements to the left are worse than the placements to the right (Amirkhanov et al. 2010)

Synchrotron tomography (s-XCT) typically employs a parallel beam geometry. As compared to X-ray tubes for laboratory XCT systems, synchrotron radiation offers a significant advantage by its nearly parallel beam of high brilliance. Synchrotron sources deliver a very high flux which is at least 1000 times larger than X-ray tubes at small source size (Requena et al. 2009; Kastner et al. 2010a). The spatial resolution can be increased by applying monochromatic radiation, which also allows the generation of phase contrast of internal interfaces between phases. The X-ray beam is practically parallel and there is no magnification of the image on the detector as shown in Fig. 3. The parallel, monochromatic beam set-up allows to generate an exact and quantitative reconstruction, free of geometrical and beam hardening artifacts. Synchrotron tomography itself is known since the 1990s (Kastner et al. 2010a; Harrer and Kastner 2011). The high-quality imaging capabilities of synchrotron radiation have led to a number of facilities all over the world (e.g., ESRF (European Synchrotron Radiation Facility) in Grenoble, Swiss Light Source (SLS) in Villigen, Hasylab in Hamburg, Elettra Sincrotrone Trieste, National Synchrotron Light Source (NSLS) and Advanced Photon Source (APS) in the USA). Some of these facilities have stations especially dedicated to high-resolution XCT that are available to general users from

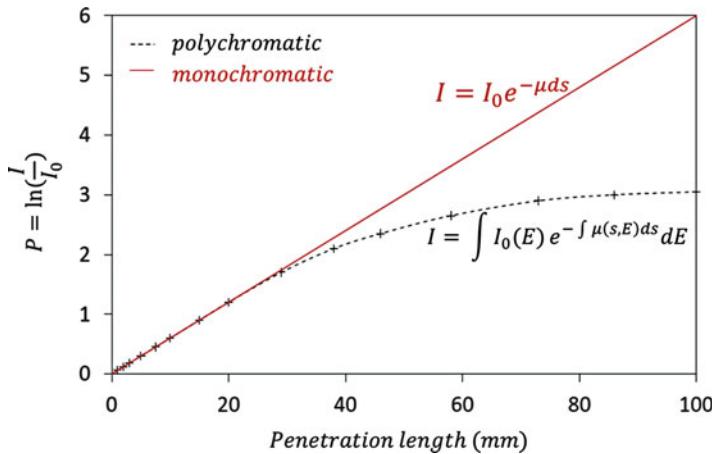


**Fig. 3** Fan beam geometry: micro-focus X-ray source with 1D line detector array: the sample requires vertical translation to be scanned (top). Parallel beam geometry: synchrotron sources ensure parallel X-ray beam. The sample is not magnified on the detector (bottom)

outside. It should be noted that s-XCT is limited to relatively small specimens, typically 5–10 mm to realize the advantages of synchrotron X-ray sources. s-XCT can reach resolutions well below 1  $\mu\text{m}$ . Spatial resolution of 40 nm has been achieved on ID22 by applying KB mirrors recently at ESRF (Requena et al. 2009).

## Artifacts

One of the major problems of cone beam XCT is artifacts. Artifacts are artificial structures in the resulting XCT dataset which do not correspond to a real sample feature. They are components of image discrepancies which are not explained by noise or transmission properties and which lead to problems in measurement interpretation and dimensional accuracy. Cone beam XCTs are prone to artifacts such as beam hardening, scattered radiation, streaking artifacts, partial volume, ring artifacts, and aliasing-induced artifacts. In Fig. 4 some of the most common artifact types in XCT are depicted together with their physical background.



**Fig. 4** Mono- and polychromatic radiation curves.  $P$  is a projection value obtained by normalizing Eq. (7). (Adapted from Kasperl 2005)

Beam hardening is one of the most prominent artifacts, which can be explained in the following way: Lambert-Beer's law for a mono-energetic X-ray spectrum can be expressed as follows:

$$I_{\text{mono}}(E) = I_0(E) \cdot e^{-\int_0^s \mu(\rho, Z, E) \cdot ds} \quad (7)$$

$\rho$  is the density of the scanned material and  $Z$  the atomic number. Emitted photons can reach an energy between zero and  $E_{\max} = U_b \cdot e$ .  $U_b$  is the maximum X-ray tube voltage and  $e$  the elementary charge. For the respective spectrum the number of emitted photons per energy level is not constant. X-ray tubes generate a polychromatic X-ray spectrum  $S(E)$ . Additionally, the detector used has an energy-dependent efficiency  $D(E)$ . The energy dependencies of the tube spectrum and the detector efficiency cannot be obtained separately. Therefore, their product  $S(E) * D(E)$  can be understood as a weighting factor for each present energy. This results in a modification of Lambert-Beer's law:

$$I_{\text{poly}}(E) = I_0 \int_{E_{\min}}^{E_{\max}} S(E) \cdot D(E) \cdot e^{-\int_0^s \mu(\rho, Z, E) \cdot ds} dE \quad (8)$$

If an X-ray beam with a broad-band energy spectrum passes through an object, the spectrum changes along the path. Usually, the low energy (soft energy) beams are more strongly absorbed than the high-energy (hard energy) beams. Therefore, the polychromatic X-ray beam becomes harder as it passes through matter. This leads to the so-called beam hardening artifacts or cupping effect (Kasperl 2005).

**Scattered radiation** plays an important role for materials with high attenuation coefficients. Scattering occurs when the incident X-ray photon is deflected from its

original path by interaction with matter. These scattered X-rays may become a considerable part of the XCT signal resulting in artifacts. One correction strategy for scattered radiation was introduced by Joseph and Spital (1982); however this method works only for simple object geometries. Methods which consider the exact geometries (e.g., ray-tracing methods or Monte Carlo simulations) are computationally intensive and unacceptable for industrial applications to date (Harrer and Kastner 2011).

**Reconstruction artifacts:** For the Feldkamp reconstruction algorithm, respective Feldkamp artifacts occur, especially for larger cone beam angles and at surfaces of the specimens perpendicular or almost perpendicular to the rotation axis. According to Tuy-Smith's sufficiency condition (Tuy-Smith 1983) an exact reconstruction without losses of spatial resolution using a circular acquisition trajectory is possible, if all surfaces intersecting the object intersect the trajectory of the X-ray source at least once. This means for cone beam imaging, each plane crossing the object must intersect the trajectory of the focal point at least once. This condition is only true for the central XY plane of the measuring field using standard cone beam XCT. Losses of spatial resolution in the direction of the rotation axis can lead to rhombus-shaped artifacts (i.e., Feldkamp artifacts).

Other important artifacts are:

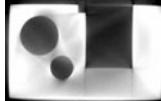
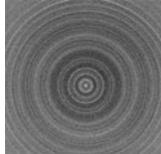
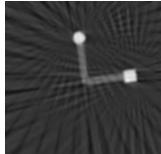
- **Ring artifacts:** result from sensitivity drifts of the individual detector pixels or faulty pixels. As shown in Table 3, they appear as concentric rings. A possibility for their reduction is a random, horizontal movement of the detector between the acquisition of each projection image. Ring artifacts may be reduced or compensated to a high degree.
- **Partial volume artifacts:** a specimen typically features a sharp well-defined boundary. Due to the limited resolution of the detector, this boundary will not

**Table 3** Typical types of artifacts in XCT together with the physical reasons

Artifact	Sample Image	Physical background
Beam hardening		The incident radiation is polychromatic. The energy-dependent absorption shifts the spectral center of mass to higher energies. As the material scanned is homogeneous, the same gray value should be visible throughout the solid material region. Drill holes should show the same gray value as the surrounding air. This artifact also known as cupping
Scattered radiation		Scattered radiation originates through the physical effect of Compton and other scattering processes. The edges of the scanned homogeneous material are blurred and contrast is deteriorated
Streaking artifacts		Streaking artifacts originate through beam hardening. High-density objects may either not be fully penetrated by the incident X-ray radiation or absorb the lower energy parts of the spectrum. Furthermore, another reason causing streaking artifacts is limited detector dynamics

(continued)

**Table 3** (continued)

Artifact	Sample Image	Physical background
Cone beam artifacts		Cone beam or Feldkamp artifacts typically occur for planar faces of the specimen, especially in the upper and lower regions of the specimen. Using the Feldkamp reconstruction algorithm in circular cone beam geometry, the backprojection process suggested by Feldkamp is identical to the inverse Radon transform only in the midplane of the cone beam. An exact reconstruction is therefore only possible in this central slice. Consequently, the image quality degrades with increasing cone angles and distance to the mid plane due to the incompleteness of the acquisition geometry
Partial volume effect		Too low spatial resolution modifies gray values especially at the edges of features. Although there are only two different materials present in the sample as well as the surrounding air, the image shows a number of different gray values at the edges of the material which shows material regions, which are only partly covering the respective pixels of the detector
Ring artifacts		Ring artifacts are introduced due to irregularities of neighboring detector pixels. The rotation of the specimen produces circular artifacts
Aliasing-induced streaks		If the number of projections is too low, especially the Feldkamp reconstruction algorithm (but also others) results in streaks. When backprojecting projection images in the reconstruction space, the streaks in the reconstructed volume will decrease if the number of projection is increased
Motion artifacts		Geometrical instability and shifts lead to blurring

necessarily be located exactly between two detector elements. The intensity of the X-ray on these two detector elements will be averaged and the object is blurred. Reducing the voxel size decreases this effect.

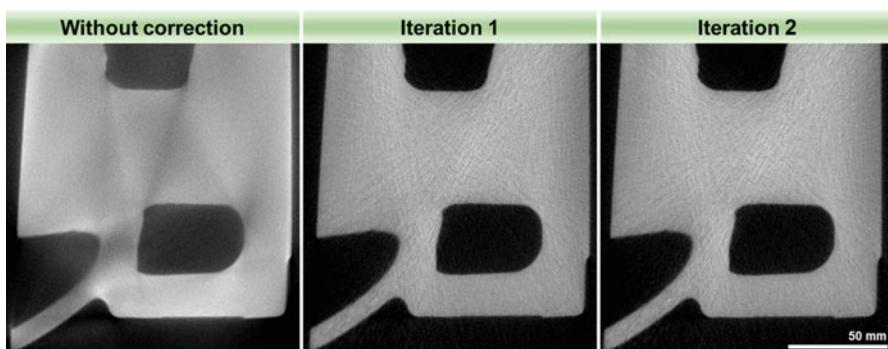
- **Aliasing-induced artifacts:** when the number of projections is too low, such artifacts appear.
- **Motion artifacts:** due to geometrical instability and shifts which lead to blurring in the data.
- **Afterglow artifacts:** XCT results may deteriorate due to a long glow time of the X-ray detector.

## Artifact Correction Methods

**Beam hardening reduction:** A widely used beam hardening correction procedure for mono-material components is linearization, which was formulated by Herman (1979). The principle of linearization consists of transforming the measured polychromatic projection values into corresponding monochromatic values. In practice, a nonlinear characteristic correction curve is experimentally retrieved from projection images of homogeneous reference samples (e.g., step wedge) with well-known geometries. The reference needs to be composed of the same material as the specimen. This correction curve is applied to the projection images as a pre-processing step before reconstruction starts based on Eq. (7). A typical correction curve is shown in Fig. 4.

An iterative algorithm to reduce scattered radiation and beam hardening in cone beam XCT called iterative artifact reduction (IAR) was proposed by Kasperl (2005). Projection images are pre-processed using the correction curve, which is extracted from a post-processing step of reconstructed data. Therefore, the correction curve is enhanced in each iteration and consequently the quality of the dataset is improved. The major disadvantage of this method is found in the fact, that for each specimen and material a new characteristic curve needs to be determined and a reconstruction per iteration is necessary. Furthermore, the quality of the correction curve depends on the segmentation quality of the considered material, and typically additional noise is introduced in the beam hardening reduced XCT dataset. Figure 5 shows the result of an iterative artifact reduction applied on XCT data of an aluminum sample.

A multi-material correction method by Krumm et al. (2008) expands the IAR method for multi-material components: The materials in the reconstructed XCT image are initially segmented. Then mono-energetic and poly-energetic reprojections are computed using ray casting. The difference between both reprojections is used as a correction value for the initial projections. The whole correction process is iterative. One important limitation of the iterative methods is a



**Fig. 5** XCT slices of aluminum cylinder head scanned without correction and IAR-corrected with one and two iterations. In the IAR-corrected data, the measurement artifacts are significantly reduced, but additional noise is introduced

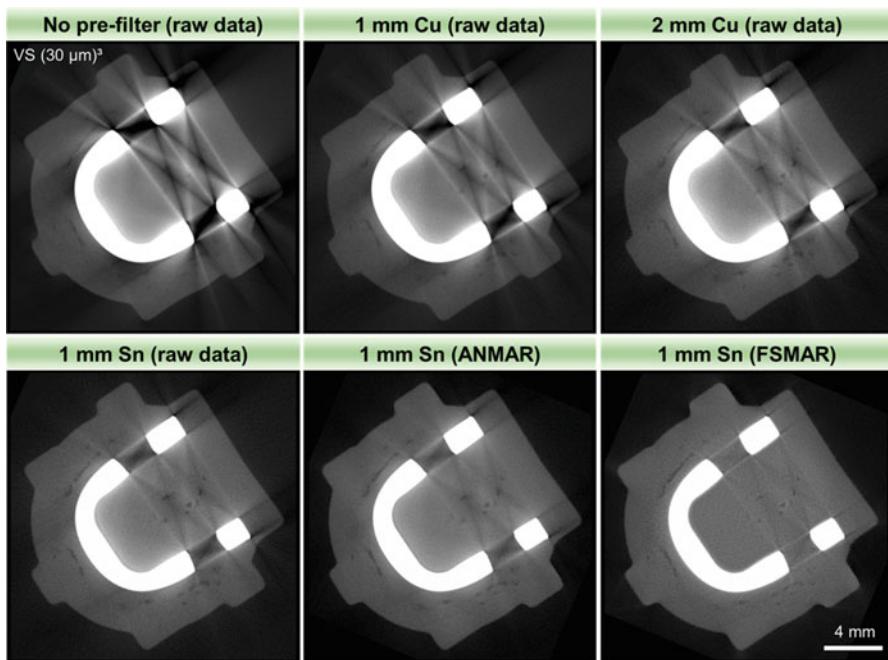
high computational cost which results in long processing times. In many cases such a performance is not suitable for XCT where high-resolution 3D data are used.

**Metal artifact reduction:** There is a large variety of metal artifact reduction (MAR) methods based on sinogram inpainting, which have been used first in the field of medical imaging (Meyer et al. 2010). These methods are especially beneficial for multimaterial-components like plastic housings with metallic inlets or plugs with metallic pins. MAR methods rely on the detection and deletion of metal traces within the sinogram. A forward projection of the segmented metal parts helps to identify metal areas in the sinogram. Then, the metal data in the sinogram is deleted in all affected areas, and a subsequent data completion of missing data via interpolation is done. The interpolation step may cause blurring of edges due to information loss of deleted metal traces as well as the formation of streaks. Furthermore, inpainting techniques typically fail if the metal to be segmented is hollow. In this case the interior hollow regions will be lost. Meyer et al. have proposed additional modifications based on an improved interpolation (Meyer et al. 2010). These more sophisticated MAR-techniques are listed below:

- Normalized MAR is based on MAR with an additional normalization step of the projections before and after interpolation by calculating a prior image of the matrix material via a segmentation step of the initial image. The original sinogram is then divided by a forward projected prior image before interpolation, leading to a flat sinogram which can be interpolated in a smoother way.
- Adaptive normalized MAR is based on normalized MAR with a weighting sum of the original projections.
- Frequency Split MAR is based on normalized MAR with an additional weighting approach of a combination of high- and low-pass filtered original.

As an example, we present results of an injection-molded part with steel insert (Gusenbauer et al. 2016). The steel insert is causing severe artifacts in XCT scans when no additional pre-filtering is applied to the X-rays before penetrating the specimen (see Fig. 6, top left image). For steel these artifacts are much stronger as compared to aluminum. The metal artifacts, visible as bright and dark streaks, can be reduced quite significantly by adding additional tin pre-filters. A targeted SNR value can be only kept constant at the expense of additional measurement time. The application of the ANMAR method (see Fig. 6, bottom center) shows improvements in terms of artifact reduction, since in specific projection images, less metal has to be penetrated. At these particular angular positions, the original projection images are considered with a higher weighting factor as the others in the weighted sum of original and NMAR corrected projection images. FSMAR in Fig. 6 bottom right image combines the high frequencies of the uncorrected image, where all available data are used for the reconstruction, with the more reliable low frequencies of an image that was corrected with the inpainting-based NMAR method. FSMAR indicates the best image quality in terms of observable streaking artifacts.

**Methods to correct scattering and ring artifacts:** For the reduction of scattering artifacts there are several methods known and used. Scatter reduction techniques can



**Fig. 6** XCT slice images of an injection molded part with steel insert. Original data (top left) versus 1 mm Cu, 2 mm Cu, as well as 1 mm Sn prefiltered data with ANMAR and FSMAR, voxel size ( $30 \mu\text{m}$ )<sup>3</sup>. Adapted from Gusenbauer et al. 2016)

be grouped in scatter reduction techniques using anti-scatter grids and scatter correction techniques (Wiegert 2007). Anti-scatter grids consist of a multitude of lead strips, separated by an interspace material of very low atomic number such as carbon fiber or aluminum. Placing such grids close to the detector causes a reduction of scattering within the grid, only letting pass X-rays that are parallel to lead strips. In contrast, scatter correction techniques remove the scatter by first estimating the scatter and then subtracting its share from each projection.

For reducing ring artifacts several methods are available for XCT. The most effective way of avoiding ring artifacts is to use hardware correction methods, which eliminate artifacts from projections (Carmignato et al. 2018). Ring artifacts may be removed by slightly moving the detector (or the specimen) during data acquisition. Consequently, the characteristics of detector elements are averaged, reducing the occurrence of ring artifacts. However, movements of the detector or the specimen are time-consuming and might integrate undesired imprecisions due to positioning errors. Whenever hardware corrections cannot be used, software methods may be used. A typical pre-processing method is the flat-field correction. Post-processing approaches aim to eliminate ring artifacts by means of filters such as Fourier filters, wavelet filters, combined wavelet-Fourier filters, and median filters. These filters are typically used to remove ring artifacts from the sinogram or from the reconstructed XCT data.

## Further Important XCT Methods

There are a large number of other XCT methods which use different acquisition trajectories and different reconstruction techniques which you can find in literature. In the following sections, a limited selection of the most relevant ones for NDE is presented:

**Multi-scale tomography:** The specifications of typical XCT systems (X-ray tube and voxel sizes) are compared in Table 4.

The different size scales of the features of interest and of the respective material volume require appropriate tomographic methods: macro-XCT or standard XCT ( $>10\text{ }\mu\text{m}$ ); micro-focus X-ray tomographs,  $\mu$ -XCT (currently  $>3\text{ }\mu\text{m}$ ); sub- $\mu\text{m}$  XCT/ nano-XCT (currently  $>0.4\text{ }\mu\text{m}$ ); and synchrotron tomography s-XCT (currently  $>0.2\text{ }\mu\text{m}$ ) (Harrer and Kastner 2011). s-XCT applying KB mirrors have reached resolutions down to  $0.04\text{ }\mu\text{m}^3$  at ESRF/ID2237 (Requena et al. 2009). Figure 7 shows dimensional scales for voxel size and object size (diameter) for various XCT methods. XXL-XCT/robot-based XCT, macro-XCT,  $\mu$ -XCT, and nano-XCT systems cover a range of more than four orders of magnitude for both parameters.

**Nano-cone beam XCT and nano-XCT with X-ray optics/X-ray microscopy CT:** Laboratory Sub-micron/Nano-XCT with cone beam geometry (projection XCT) without using X-ray optics can reach resolutions down to  $0.3\text{--}0.5\text{ }\mu\text{m}$  (Kastner et al. 2017a). The resolution is mainly limited by the limitations of X-ray tubes where the X-ray flux

**Table 4** Technical data of various X-ray computed tomography systems

XCT device	X-ray source	Best resolution [ $\mu\text{m}$ ]	Min. voxel size [ $\text{nm}^3$ ]	Max. Sample size [mm]
Standard Medical XCT <sup>a</sup>	100–140 kV tube	300	—	Size of a human
Rayscan 250E <sup>b</sup>	225 kV micro- and 450 kV macro-focus tube	~10	5000	Diameter: 600 Length: 1800
GE phoenix nanotom m <sup>c</sup>	Nano-focus tube 180 kV	~1.0	300	Diameter: 210 Length: 250
RX Solutions Easytom 160 <sup>d</sup>	Nano-focus tube 160 kV	~0.5	50	Diameter: 200 Length: 700
Zeiss Xradia 810 Ultra (X-ray microscope) <sup>e</sup>	35 kV + X-ray lenses	~0.05	16	0.016 (field of view) or 0.065 at 64 nm voxel size
Synchrotron XCT: ESRF Grenoble, e.g., ID19 or ID22 <sup>f</sup>	6–80 kV	0.28, with KB mirror down to 0.04	—	Typically, 5–10

<sup>a</sup>XCT suppliers: GE Healthcare: Philips, Siemens, Toshiba, etc.

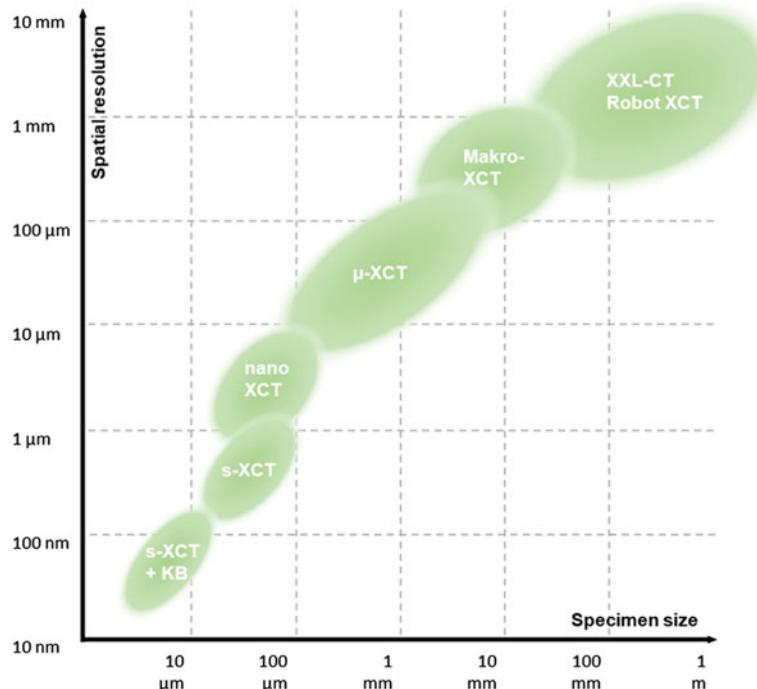
<sup>b</sup><http://www.rayscan.eu/>

<sup>c</sup>[https://www.gemeasurement.com/sites/gemc.dev/files/geit-31344en\\_nanotom\\_m\\_0517.pdf](https://www.gemeasurement.com/sites/gemc.dev/files/geit-31344en_nanotom_m_0517.pdf)

<sup>d</sup><http://www.rxsolutions.fr/>

<sup>e</sup>[www.xradia.com/zeiss-xradia-810-ultra](http://www.xradia.com/zeiss-xradia-810-ultra)

<sup>f</sup><http://www.esrf.eu/home/UsersAndScience/Experiments/StructMaterials/ID19.html>



**Fig. 7** Multi-scale X-ray tomography showing the different spatial resolution plotted against the respective specimen sizes accessible by cone beam lab-based XCT systems (XXL-XCT/robot-based XCT, macro-XCT,  $\mu$ -XCT, nano-XCT/sub- $\mu$ m XCT), as well as s-XCT

dramatically decreases with decreasing focal spot size. Nevertheless, novel ultrabright tubes with high flux can provide reasonable X-ray flux for focal spot sizes below 1  $\mu$ m. These kinds of tubes often use (structured) tungsten on diamond transmission targets since diamond features high thermal conductivity and the material of choice for the filament is often LaB6. In this way a minimum focal spot size of 0.5  $\mu$ m for X-ray tubes and a physical resolution of around 0.5  $\mu$ m for a cone beam XCT system can be reached (see Table 4). XCT data of a nano-XCT system are presented in Figs. 11 and 12.

A possible solution to overcome the limitations of high-resolution cone beam XCT is found in using X-ray optics (Fresnel zone plates or multi-layer Laue lenses) for XCT, which can also be called transmission X-ray microscopy. Here, a high flux X-ray tube is used. The X-rays are energy filtered and collimated and focused by the first X-ray lens on the specimen to reach high resolutions. After the penetration of the specimen there is another X-ray lens (objective) widening the X-ray beam onto the detector. XCT systems with X-ray optics can reach resolutions below 100 nm (Kastner et al. 2017a). However, there are several severe limitations. First, due to the X-ray optics, the system is restricted to a specific design energy, and second, there is significant intensity loss at the collimator and at the lenses. Therefore, the scanning times for high-resolution XCT scans using X-ray optics are extremely long (e.g., 24 h or more).

**Helical XCT:** Conventional cone beam XCT uses circular scanning trajectories, even though not all the required information for an artifact-free reconstruction of the object can be provided. According to the Tuy-Smith sufficiency condition (Tuy-Smith 1983) losses of spatial resolution in the direction of the rotation axis using circular cone beam CT can lead to Feldkamp reconstruction artifacts. Helical XCT (also called spiral computed tomography) is a method using helical data acquisition trajectories (Carmignato et al. 2018). During data acquisition the specimen is also shifted along the rotational axis. Thus, from the specimen's perspective, the source of radiation is moving on a helical trajectory. For the data acquisition using helical trajectories, the Tuy-Smith sufficiency condition is fulfilled for all XY planes. For this reason, the specimen may be reconstructed exactly, and cone beam artifacts can be avoided. An additional advantage is that components with unrestricted lengths may be scanned.

**Region-of-interest XCT (ROI-CT) or local XCT:** For conventional CT, the best possible resolution is limited in terms of voxel size by the quotient of the maximum lateral object diameter and the number of detector pixels. Both of these parameters are typically fixed in conventional XCTs and thus the maximum resolution is fixed. ROI-CT overcomes this limitation by focusing and scanning a region of interest of the specimen in higher resolution. For this purpose, special reconstruction algorithms are required. However, these algorithms do not work perfectly. Depending on the measurement task and reconstruction methods used, artifacts occur.

**Limited-angle XCT:** Limited-angle reconstruction is a typical requirement in industrial XCT imaging, especially when scanning planar objects, which do not allow a conventional circular scan trajectory. An XCT scan for planar specimens is often prohibited or impossible due to collision risks and the limited penetration capability of the X-ray beam. Therefore, only penetration images in a limited range of angles may be acquired, resulting in a limited-angle XCT reconstruction problem. A general solution to the exact reconstruction from limited-angle projections is not possible as Tuy-Smith's sufficiency condition is violated. In addition to missing projections, the limited number of projections will introduce severe streaking artifacts in the reconstruction. The most common approach to reconstruct an object from limited-angle data is to complete the set of projections adding empty images for missing projections to be able to apply the classical FKD reconstruction algorithm. However, the result will suffer from strong artifacts. Alternatively, a priori data may be incorporated into the reconstruction process to reduce artifacts. Concepts in this respect either employ data sparsity or existing data from CAD-models (Pereira et al. 2017; Ueda et al. 2017).

**Computed laminography: (CL)** When large specimens are to be inspected with high resolution, XCT reaches its limitations. This is especially true for planar objects. XCT requires the specimens to be imaged from many positions around a circular trajectory. This is not possible for extended objects either due to geometrical restrictions or because of the high absorption of the X-rays along the longitudinal directions. In these cases, digital X-ray computed laminography (CL) provides a viable alternative to XCT (Zhou et al. 1996). Specimens are scanned using an X-ray source and detector that synchronously move in opposite

directions in planes above and below the specimen. Only the features in the specimen that lie on the focal plane are projected on the same location of the detector, whereas all other structures are out of focus and are therefore blurred in the projection image during the relative motion of the system. As the X-ray source and the detector only travels in planes above and below the object, the specimen is irradiated from a limited angular range only. For this reason, laminography reconstructions suffer from a limited depth resolution (Hanke et al. 2008; Carmignato et al. 2018). For reconstruction, tomosynthesis and algebraic reconstruction techniques (ART) are used. A variant of linear sampling laminography is the swing laminography also called limited-angle tomography. This scanning geometry can be realized with standard CT scanners. The object is not rotated by full 360° but only by e.g., ± 30°. The data can be reconstructed by limited-angle filtered backprojection or by algebraic reconstruction techniques (Hanke et al. 2008).

**XXL-CT and robot XCT:** XCT is not only used for small parts. There is an increasing demand to scan large parts such as containers, whole cars, and airplane wings. For such applications the geometrical limitations and limitations in X-ray penetration need to be overcome. Therefore, XCT systems are equipped with linear accelerators featuring X-ray energies up to 10 MeV (Reims et al. 2014). The achievable image quality is limited using standard detectors due to the dominating absorption effect of Compton scattering at high energies. Most high-energy XCT applications are therefore based on line detectors, which allow shielding against scattered radiation. The use of line detectors increases the quality of imaging but with the drawback of time-consuming scans for large volumes. Such an XXL-CT setup was installed at Fraunhofer IIS, Development Center X-ray Technology EZRT (<https://www.iis.fraunhofer.de/en/fz/zfp/tech/hochenergie-computertomographie.html>). (Last accessed 30.05.2018)). Here, a 9 MeV linear accelerator is used as X-ray source, the turntable is 3 m in diameter and it can handle specimens with a weight of up to 10 t. The line detector is 4 m in length at a pixel pitch of 400 µm. With such a system, entire cars, aircraft fuselages or fins and shipping containers can be scanned with a reasonable quality. Robotic XCT systems may also be used for larger specimens (De Chiffre et al. 2014; Kastner 2016). Using this concept, X-ray source and detector are each mounted on robots, which allow mimicking arbitrary scanning trajectories around the specimen using robots (see Fig. 8). However, due to the achievable positioning accuracy of the robots, the achievable resolution of the XCT data is also limited. Conventional XCT trajectories robot-based XCT may also be used in laminographic mode for nondestructive X-ray inspection of large and bulky objects. Similar large-scale XCT systems can be found at different places in Europe, in the USA, and all over the world.

**In-line XCT/high-speed XCT:** Within the last years, XCT has become an important tool in the field of serial production. This is due to its unbeatable ability of nondestructive evaluation of inner sample structures. For in-line XCT and high-speed XCT, the acquisition and reconstruction times were reduced from hours to minutes in the case of standard geometries (Janssens et al. 2016; Pereira et al. 2017). Future optimizations of the method will result in scanning times of well below a



**Fig. 8** Robot-based XCT system: X-ray source and X-ray detectors are mounted on collaborative robots, which are mounted themselves on linear axes. The specimen is located in the center and may be scanned using arbitrary scanning trajectories. (Image reproduced with permission from FACC Operations GmbH)

minute, so that a true three-dimensional XCT testing will be available for series production. This progress is mainly possible by:

- Using the preliminary knowledge of the specimen for optimized measurement trajectories/strategies to get optimized projection data.
- Using the preliminary knowledge of the specimen for efficient defect detection and/or geometrical analysis.
- Optimization and parallelization of XCT data evaluation pipelines.
- Increasing the tube power and using better detectors.

Typical applications of in-line XCT and high-speed XCT are void detection in aluminum parts, detection of cracks and inclusions in plastic parts, mixed material assembly analysis in multi-material components, metrological analysis of different parts, etc.

**Dual-energy XCT or multi-energy XCT (DECT/MECT):** The basic idea of DECT/MECT has been known since the 1970s (Millner et al. 1979; Heinzl et al. 2007). DECT/MECT is based on the fact that material components present in a complex object can be distinguished utilizing two X-ray energy setups, low-energy and high-energy XCT, which apply different X-ray spectra on the specimen. A material decomposition using DECT/MECT exploits the fact that differences between the attenuation coefficients of materials with different atomic number significantly increase with decreasing tube potential, while at high energies the respective change is much slower. This effect may be achieved by performing two

or more CT scans with different energy spectra (low and high energy) or by employing a single photon counting detector with a spectroscopic capability. A material decomposition may be achieved in a pre-reconstruction space (projection data) or a post-reconstruction space (volume data). Using DECT/MECT two different materials with the same density but different chemical composition may be distinguished and the inspection and geometrical measurement of multi-material components like carbon fiber-reinforced plastic composite combined with metal components or electrical connectors with encapsulated metal parts may be improved.

**XCT simulation:** Simulation plays an increasingly important role in the field of XCT. Over the years its applications and objectives have become highly diverse (Reiter et al. 2016). Some of the possible objectives of XCT simulation are found below:

1. Improvement of the understanding of XCT phenomena to provide helpful information for data analysis and interpretation of the generated results.
2. Evaluation of the capabilities as well as identification of the limitations of XCT methods and systems.
3. “Virtual testing of algorithms”: XCT simulation may be used to generate datasets of virtual specimens to test and evaluate algorithms for data analysis.
4. Prediction of XCT results to identify optimal XCT scanning parameters and acquisition strategies. A simulation tool may suggest parameters to make the results more objective and generate user-independent maximum quality.

For XCT simulation, models have to be extracted from the respective components and effects. The modeling process covers the following sections of an XCT system:

1. X-ray source model (X-ray spectrum, focal spot size and stability)
2. Interactions of X-rays with matter of the specimen or a possible pre- or postfilter
3. Detector model and interaction of X-rays with the detector (detector efficiency (DE), pixel size, image lag, noise, scattering effects, interaction of X-rays with scintillator, etc.)
4. Acquisition geometry of the XCT system including static and dynamic misalignments of each component

Monte Carlo simulations and hybrid approaches are hardly applicable, because of their high calculation effort. Many software simulators for NDT and metrology applications of XCT are available, such as SimCT (<http://www.3dct.at/cms2/index.php/en/software-en/simct>. (Last accessed 16.05.2018)), Scorpions XLab (<https://www.vision.fraunhofer.de/de/technologien-anwendungen/projekte/roentgentechnik/scorpions.html>. (Last accessed 16.05.2018)), aRTist (Analytical RT Inspection Simulation Tool – BAM) ([www.artist.bam.de](http://www.artist.bam.de). (Last accessed 16.05.2018)), and CIVA computed tomography module (<http://www.extende.com/computed-tomography-with-civa>. (Last accessed 16.05.2018)). However, there is no commonly accepted standard XCT simulation tool available right now.

## Applications of X-Ray Computed XCT

### Overview of Applications

XCT has gained increasing importance and acceptance in the area of examination and inspection of industrial specimens within recent years. The applications of XCT are found in many different industrial fields (Kruth et al. 2011; Maire and Withers 2014; Kastner 2016) such as:

- Materials science
- Aerospace
- Automotive
- Electronics
- Medical/food technology
- Leisure and sporting goods industry
- Archaeology
- Security
- Science and Research
- Arts
- Biomedicine

The industrial XCT market is a steadily growing market with huge potential for novel yet unknown applications. Because of its advantages XCT is the method of choice in many application areas:

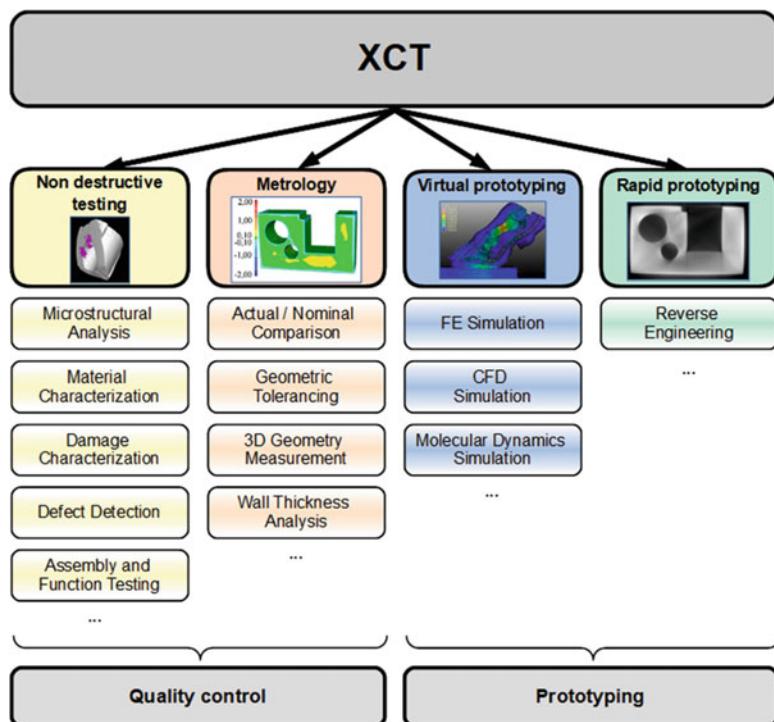
- The first and still most important application area of XCT is nondestructive testing and materials characterization. While for NDT internal structures such as shrink holes, material inclusions, or cracks are of special interest, in materials science the 3D characterization of different phases and/or filler materials are crucial. XCT supports the understanding, discovery, design of materials, which are integral parts of materials science. It also allows to evaluate the material in use in recent 4DCT setups.
- The second major application area of XCT is metrology for dimensional measurements in 3D of geometric features, mainly in the field of quality control. The primary interest in metrology is the measurement of distance, diameter, wall-thickness, angle, roundness, flatness and the 3D-geometry of inner and outer surfaces in general as well as the evaluation of their tolerances. Another important application area of XCT is reverse engineering and rapid prototyping: A computer-aided design (CAD) model is extracted from an XCT scan, which allows for reproductions and duplications of components by rapid prototyping techniques.
- XCT is also beneficial for virtual prototyping. In virtual prototyping, digital models of scanned prototypes are generated from an XCT scan. This enables improvements of the geometry of components in the development phase. It also provides data for modeling and simulation of the material in use, i.e., based on

finite element simulations, computational fluid dynamics, or molecular dynamics simulations. In addition, the modeling and computation of material properties may be performed based on XCT data, e.g., for the material design and for the design of filler media or for the design of fiber-reinforced polymeric parts.

Figure 9 presents an overview of different fields of application for the XCT technology in the industrial domain. XCT measurements may be evaluated as voxel based or surface based. Quantitative, surface-based evaluation comprises digitization and dimensional metrology. The need for traceability increases with the complexity of the task. Simple visualizations do not require absolute measures. For metrology applications traceability is crucial.

## XCT for Materials Characterization

As indicated before, the applications of XCT are widespread and highly diverse. In the following section some selected applications of XCT for materials characterization are presented in more detail.

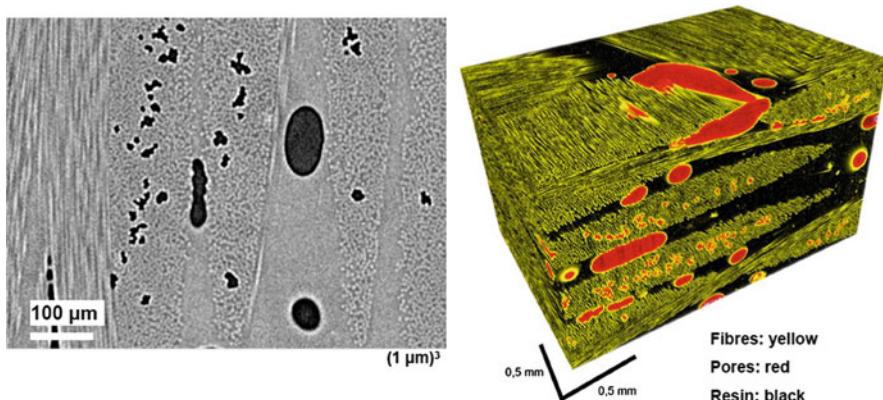


**Fig. 9** Most important fields of application for XCT

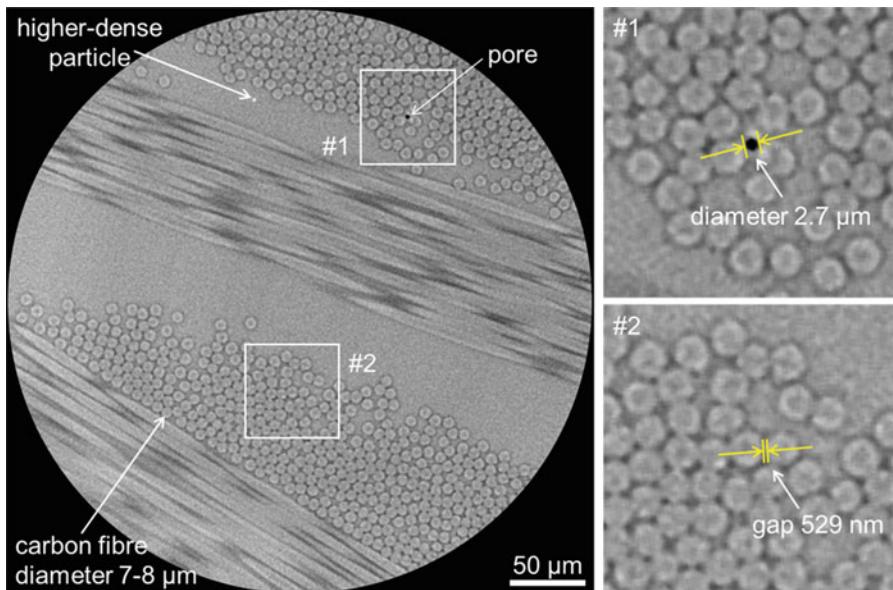
**Polymeric materials and wood:** Fiber-reinforced polymers (FRP) show an increased stiffness and increased strength to weight ratio as compared to metallic and other “traditional” materials. Therefore, the applications of these materials have become of high importance in modern industry. Nondestructive and contact-free techniques for the 3D characterization of fiber composites are increasingly demanded in process development and production. Figure 10 shows a cross-sectional XCT image and a 3D rendering of a short carbon fiber-reinforced laminate. Although the density difference between carbon fibers (density  $\sim 1.8 \text{ g/cm}^{-3}$ ) and polymer (density around  $1\text{--}1.15 \text{ g/cm}^{-3}$ ) is rather low and the diameter of the carbon fibers is only 7–8  $\mu\text{m}$ , the individual fibers can be clearly seen. The right picture of Fig. 10 shows a 3D rendering of the XCT data of a carbon fiber-reinforced polymer (CFRP) from the aeronautic industry. Carbon fibers, the polymeric matrix, and pores can be distinguished. The segmented carbon fibers are shown in yellow, the resin in black, and the pores in red.

Figure 11 shows CFRP scanned with a higher resolution (Kastner et al. 2017b). In this example, the individual carbon fibers are shown. They are parallel and perpendicular to the XCT cross section. Due to the high-resolution setup of this scan, phase-contrast effects appear at the interface between the fibers and the resin. It is also noticeable that the fibers are not completely round. Features down to a size around 0.5  $\mu\text{m}$  are visible indicating a physical resolution close to 0.5  $\mu\text{m}$ .

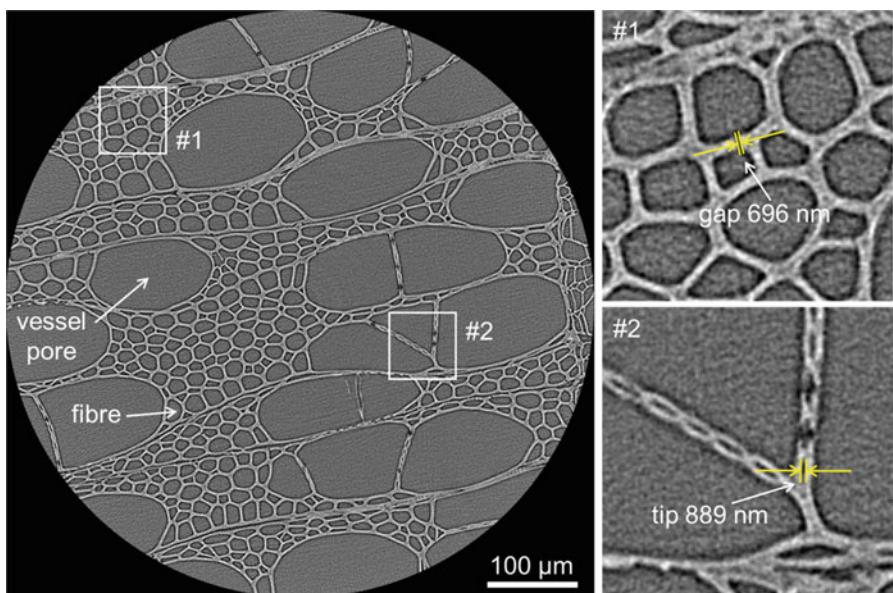
Figure 12 shows another example of a high-resolution XCT result. In this case a poplar wood sample is scanned at a voxel size of  $250 \text{ nm}^3$ . Due to phase-contrast effects between the wood cells and the air, the cellular structure of the wood and structures down to around 700 nm can be clearly seen (Kastner et al. 2017b).



**Fig. 10** Cross section of XCT data (left picture) and three-dimensional representation (right picture) of a carbon fiber-reinforced polymer sample (CFRP). Voxel size was  $1 \mu\text{m}^3$ . On the right picture, a 3D rendering of the sample is depicted. Carbon fibers are shown in yellow, the resin in black, and the pores in red (Kastner et al. 2012b)



**Fig. 11** Cross-sectional XCT pictures of a CFRP measured with a voxel size of  $250 \text{ nm}^3$ . The pictures on the right are zoomed regions of the specimen (Kastner et al. 2017b)

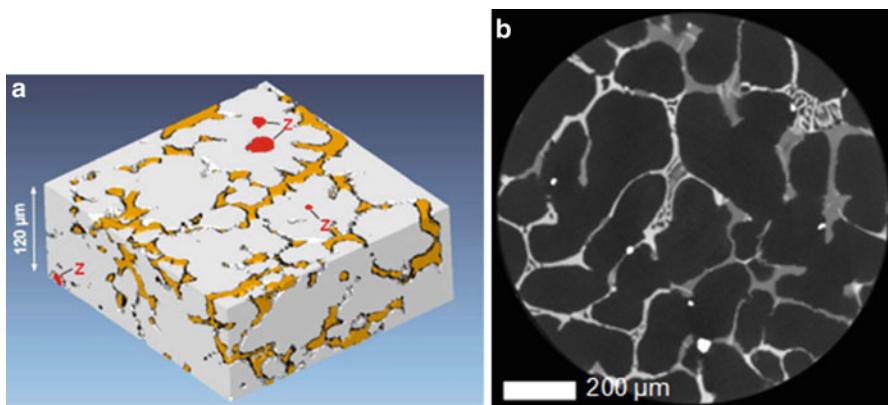


**Fig. 12** Cross-sectional XCT pictures of a poplar wood sample measured with a voxel size of  $250 \text{ nm}^3$ . The pictures on the right are zoomed regions of the specimen (Kastner et al. 2017b)

## Characterization of Metals

**Light metals:** There are many publications on XCT of light metals and other metals. As an example, the possibilities of XCT to characterize high-strength Al-Zn-Mg-Cu alloys are presented. High-strength aluminum alloys are widely used for mobility applications because of their high specific stiffness and strength, good formability, and corrosion resistance. Chill cast Al-Zn-Mg-Cu billets contain Sc and Zr, which form secondary  $\text{Al}_3(\text{Sc},\text{Zr})$  dispersoids during homogenization, but as well primary  $\text{Al}_3(\text{Sc},\text{Zr})$  in the melt. In addition, interdendritic segregations in highly absorbing elements like Zn and Cu appear. Since the interdendritic regions and the  $\text{Al}_3(\text{Sc},\text{Zr})$  particles have a similar gray value, they cannot be distinguished by simple segmentation. However, the interdendritic Cu- and Zn-rich regions and  $\text{Al}_3(\text{Sc},\text{Zr})$  particles can be differentiated by using a form factor F. This is shown in Fig. 13, where the segmented interdendritic Cu- and Zn-rich regions are visualized in orange and the  $\text{Al}_3(\text{Sc},\text{Zr})$ -particles are presented in red and marked with Z. F < 0.4 was empirically chosen for the interdendritic regions and F > 0.4 for the  $\text{Al}_3(\text{Sc},\text{Zr})$ -particles (Harrer and Kastner 2011).

The right picture of Fig. 13 shows a high-resolution cross-sectional XCT slice of as-cast MgZn6Zr1 alloy (=ZK60) modified with 2.5 wt.% rare earth metals (RE). XCT clearly shows cellular structure of the material. The material is characterized by a matrix of globular  $\alpha$ -Mg grains reinforced by an interconnected network of intermetallics along the grain boundaries. These intermetallics have higher attenuation coefficients and thus gray values than the matrix. The average dendritic cell size is in the range between 150  $\mu\text{m}$  and 200  $\mu\text{m}$ . Two distinct morphologies of intermetallic phases are observed: one is smooth and another one lamellar. The gray values of the smooth phases are significantly lower compared to the lamellar-like phases. Applying 3D analysis, it was shown that the interconnectivity of the intermetallic phases is higher than 98%.

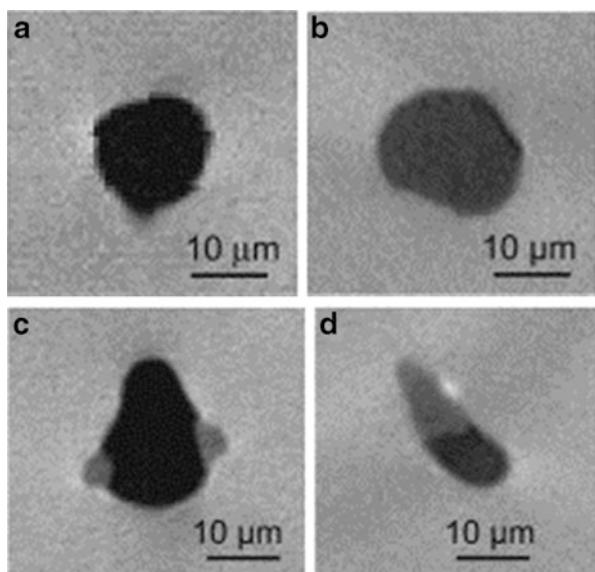


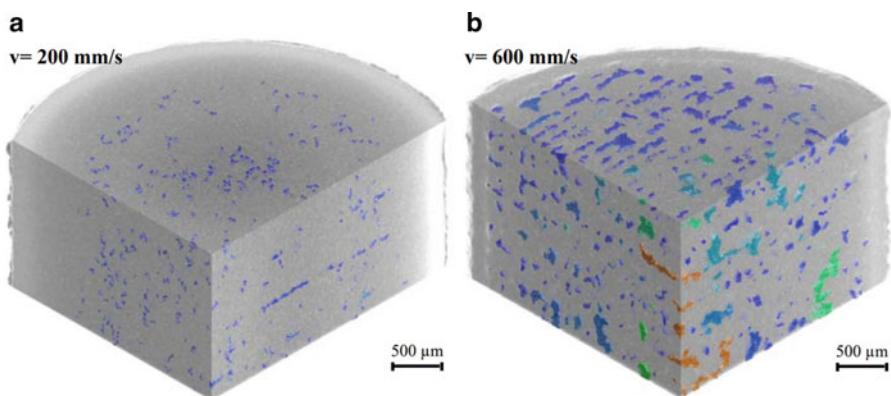
**Fig. 13** Typical arrangements of  $\alpha$ -Al dendrites (white), interdendritic eutectic regions (orange), and  $\text{Al}_3(\text{Sc},\text{Zr})$  particles (Z) in the high-strength alloy:  $\text{AlZn6Mg2Cu2}$  with Sc and Zr (a). XCT cross-sectional picture of as-cast MgZn6Zr1 alloy (=ZK60) modified with 2.5 wt.% RE (rare earth metals). Voxel size was 0.7  $\mu\text{m}^3$  (b) (Kastner et al. 2012, 2014)

**Steel:** The evaluation of XCT data of Fe-based materials is more difficult, since the respective data is usually noisy, of poor contrast, and the interpretation is affected by artifacts. Thus, inhomogeneities as shrinkage pores, gas pores, and nonmetallic inclusions are much more difficult to detect in steel, because of the higher X-ray absorption coefficient of the base metal. For instance, the spatial distribution of shrinkage pores, gas pores, and nonmetallic inclusions in steels can be characterized by XCT over the entire slab cross section. Figure 14 shows typical inhomogeneities in steel. The measurements were performed with 30 kV at ESRF-ID19 beamline with a voxel size of  $0.28 \mu\text{m}^3$ . Due to the high resolution and the monochromatic nature of the X-ray used at the synchrotron, the contrast and the resolution are high enough to distinguish nonmetallic inclusions from the pores and the steel matrix. The generated contrasts can be classified as pores, low-density inclusion, steel, and high-density inclusion. From these analyses the conclusion may be derived that the pore structure is complex, and it is not at all spherical as expected for a shrinkage pore. This is not yet possible with laboratory cone beam XCT systems. From these analyses the conclusion may be derived that the pore structure is complex, and it is not at all spherical as expected for a shrinkage pore. In Fig. 14d, a nonmetallic multiphase inhomogeneity with a diameter around  $50 \mu\text{m}$  can be seen. It features a pore and two inclusions with a lower and a higher density than the steel matrix. It consists of different nonmetallic alloys like Ca-aluminate and different spinels ( $\text{MgO}$ ,  $\text{Al}_2\text{O}_3$ ) as determined by scanning electron microscopy using X-ray fluorescence.

**Additive manufacturing: 3D-printed metals – selective laser-sintered metals:** There is a broad variety of different additive manufacturing methods for metals. Selective laser sintering (SLS) is developed to fabricate complex geometries directly from CAD data. In this additive manufacturing process, the laser selectively melts

**Fig. 14** sXCT slice of a voids pores in steel. Though the different contrast levels detected in the sXCT scan (a) pore, (b) low-density inclusion, (c) pore and low-density inclusion, (d) pore surrounded by a low-density and a high-density inclusion may be detected. The measurement was performed with 30 kV at ESRF-ID19 Grenoble with a voxel size of  $0.28 \mu\text{m}^3$  (Kastner et al. 2010a)





**Fig. 15** Three-dimensional XCT presentation of pores (black) in selective laser-sintered steel powder, manufactured with different laser scanning speeds: (a)  $v = 200 \text{ mm/s}$ , porosity = 0.9 vol.%. (b)  $v = 600 \text{ mm/s}$ , 7.8 vol.%. The distance ( $D$ ) of laser lines and alternate movements in layers ( $M$ ) of the laser are marked. Voxel size was  $7.5 \mu\text{m}^3$  (Plank et al. 2009)

powdered material by scanning across the surface of a powder bed. The choice of the processing parameters has significant influence on the quality (e.g., porosity) of the components. XCT is a powerful tool to characterize the quality of parts and to optimize manufacturing parameters. Cylindrical samples were produced with different fabrication parameters by SLS layer by layer from steel powder (X2CrNiMo17-12-2). XCT scans were performed with a voxel size of  $7.5 \mu\text{m}^3$  and the porosity values were determined for pores  $>45 \mu\text{m}$  (see Fig. 15). This sample on the left image was manufactured with a laser scanning speed of 200 mm/s and the porosity was below 0.9 vol.% (Plank et al. 2009). Tripling the laser scanning speed leads to higher porosity values (7.8 vol.%) as seen in Fig. 15 (right image). According to the pores' orientation the laser paths may be reconstructed.

## XCT for Metrology

Besides nondestructive testing the second main application area of XCT is metrology. The primary interest in metrology is the dimensional measurement of inspection features such as crucial distances, wall thicknesses, diameters, or other geometrical features and tolerances. Coordinate measuring machines (CMMs) based on tactile or optical sensors are the most common equipment for 3D dimensional measurement, because of their accuracy and flexibility. XCT is an alternative to conventional CMMs as it not just obtains surface information but also retrieves data on internal structures. XCT can generate accurate information which are not accessible to tactile or optical measuring systems (Müller 2012; De Chiffre et al. 2014; Carmignato et al. 2018). However, the inspection of geometric tolerances is a complex task, as geometric errors are related to three-dimensional features and estimating this type of error is usually based on a cloud of points that has to be measured on the machined

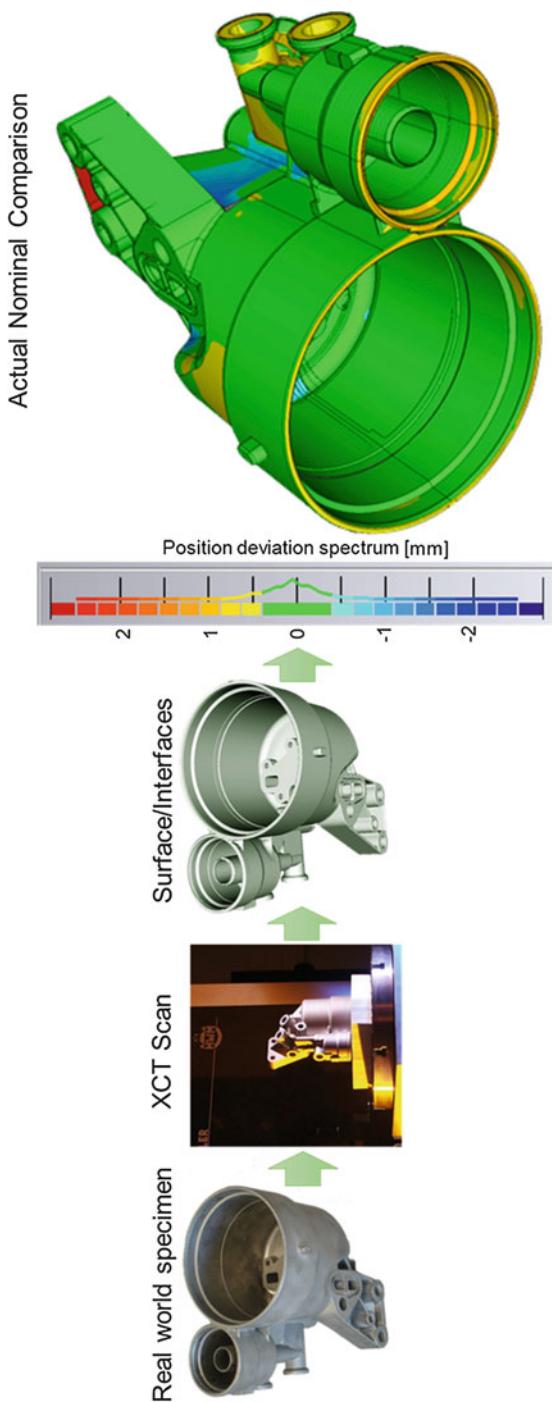
surface. XCT is of high benefit for metrological purposes where a lot of measurement features have to be inspected. Using XCT the inspection of quality-relevant features such as diameters, distances, angles, roundness, evenness, etc. can be performed within the extracted 3D model. In contrast to CMM the measurement time of XCT does not depend on the number of features to be measured. Furthermore, XCT even allows the evaluation partly or fully assembled components. Finally, the speed of metrology using XCT is an important advantage: Typical overall analysis times for industrial components like injection-molded or light metal parts including the set-up of the workpiece, the scanning, and the feature extraction are 30–60 min. For special parts and purposes, times below minutes are possible. The metrological applications of XCT are in the following different fields:

- 3D-geometry measurement
- Measurement of hidden structures and inner geometries
- Measurement of wall thicknesses
- Actual/nominal comparison which is especially important for first article inspection
- Generation of STL-representations (stereo lithography) of the surface of the specimen
- Generation of CAD-files of the component (reverse engineering)
- XCT as basis for duplication by rapid prototyping methods

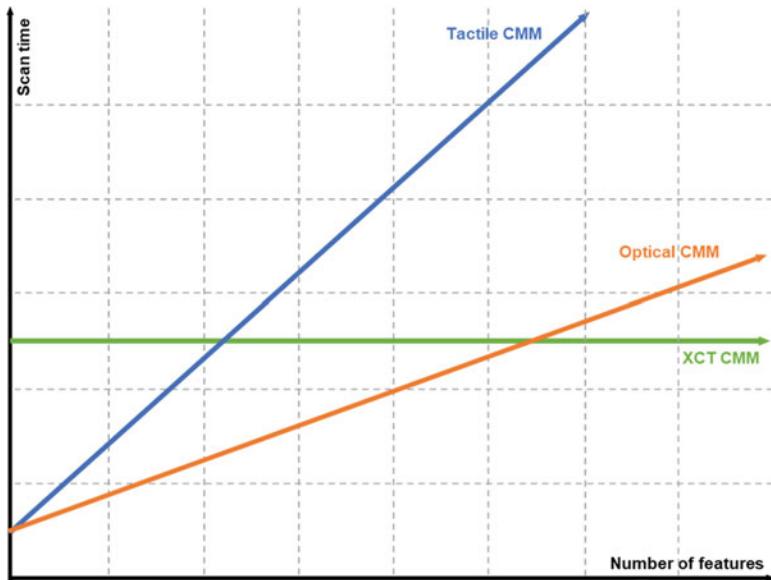
A common technique for the visualization of geometrical deviations using color coding is actual/nominal comparison: An actual/nominal comparison and the respective data processing pipeline are shown for an Al die-casting part in Fig. 16. The specimen is scanned with the XCT device. Using surface extraction techniques, the interfaces of this material toward air or the other materials present are computed and compared with a reference CAD model of the specimen. The color mapping denominates the deviations between the surface extracted from XCT data and the reference CAD data. Each position of the reference CAD model is thus coded with a color corresponding to the local deviation. Similarly, wall thickness measurement by XCT may be carried out. Here, the colors correspond to wall thicknesses computed by evaluating the thickness of the material in surface normal direction.

XCT coordinate measurements integrate a paradigm change in the way measurements occur. Whereas conventional coordinate measurements are performed directly on the workpiece surface, dimensional measurements using XCT are performed on a virtual model of the workpiece. This means that a large number of features can be analyzed at the same time without compromising the scanning time. Figure 17 shows the relationship between the scanning and the number of analyzed features for different measuring devices. It shows that the scanning time of an XCT device does not depend on the number features to be analyzed.

Although the usage of XCT for metrology is strongly increasing, XCT is not yet established as a standard measurement method in metrology. One requirement for this to happen is traceability and the knowledge of the underlying measurement uncertainty. Because of the large number of factors which influence the overall



**Fig. 16** Actual/nominal comparison between CT-data and CAD model (colors correspond to the local deviation in mm) for a die-casted aluminum part. The colors correspond to the local deviations in mm (a). Wall thickness measurement by XCT. The colors correspond to wall thicknesses in mm (b)



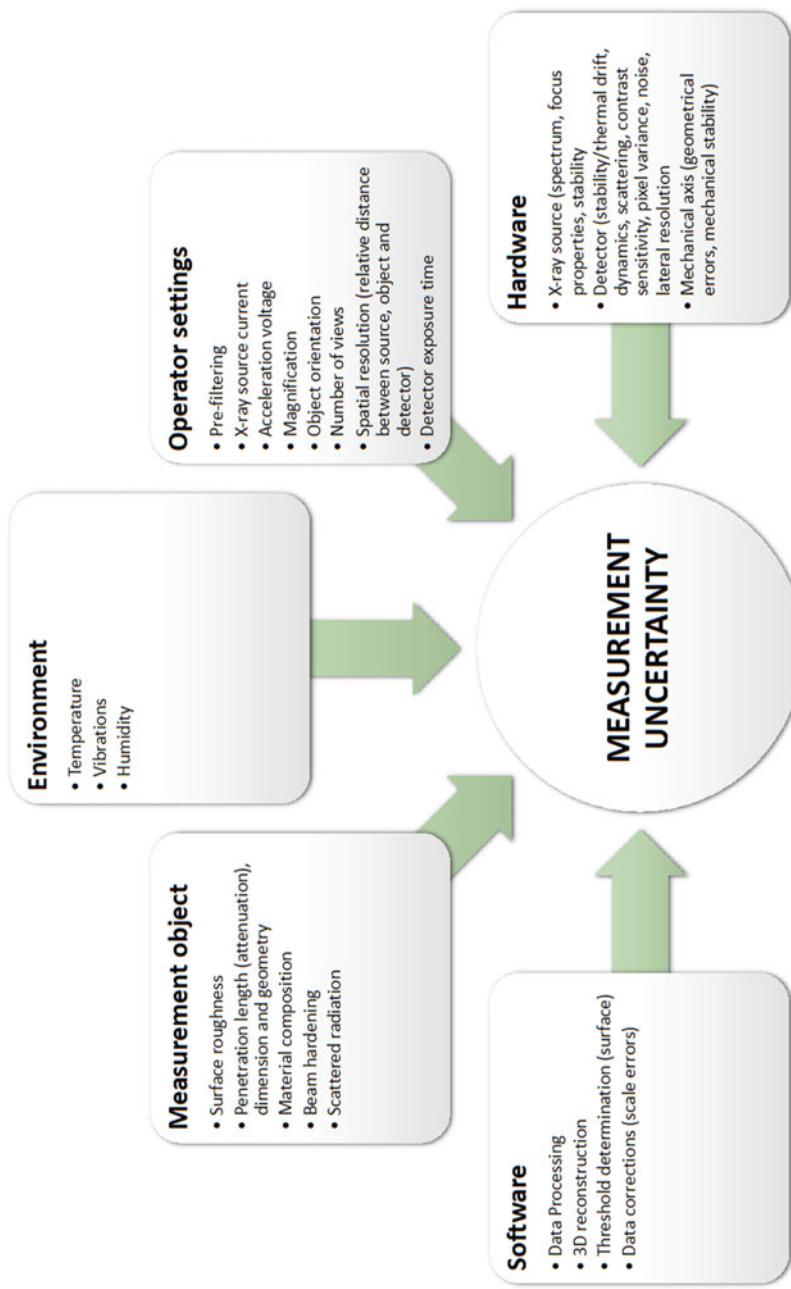
**Fig. 17** Scanning time versus the number of features to be analyzed for tactile and optical coordinate measuring machines (CMM) and XCT. (Adapted from Carmignato et al. 2018)

performance of XCT it is difficult to achieve traceability of XCT. The influence factors are described in the guideline VDI/VDE 2630-1.2 (Bartscher et al. 2018) and categorized into five groups (See Fig. 18):

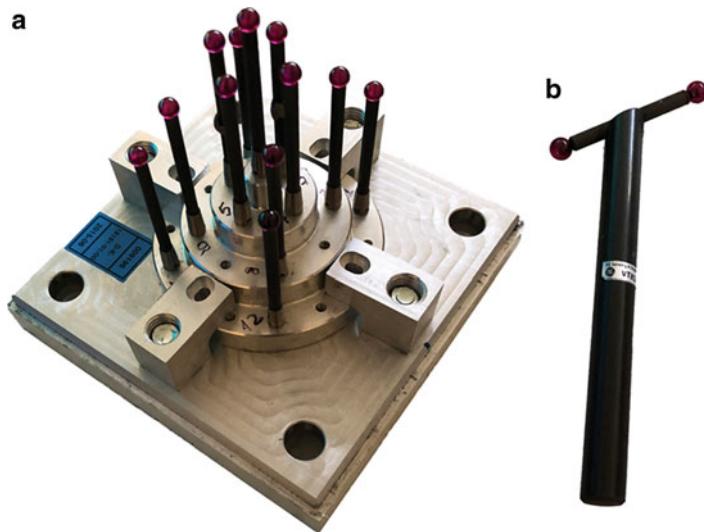
1. Factors connected with the hardware (X-ray source, rotary table, X-ray detector) and adjustment of the XCT device: for this task various reference standards such as a ball bar, a ball plate featuring spherical calottes, a calotte cube, or other parts can be used.
2. Software and data processing (3D reconstruction and reconstruction parameters, XCT data preprocessing, XCT artifact correction, segmentation and surface generation, extraction of measurement features).
3. Environment (temperature, humidity).
4. Measured object (geometry, material).
5. Operator settings (scanning parameters): more than 40 quantities are relevant.

The XCT measurement process is influenced by the X-ray source, the specimen, the detector, the scanning geometry, the environment, the user, the reconstruction, and the data evaluation. Every parameter of the whole process chain as listed below has a significant influence and needs to be aligned appropriately and accurately:

- Calibration/adjustment of the XCT device: for this task various reference standards such as a ball bar, a ball plate featuring spherical calottes, a calotte cube, or parts as shown in Fig. 19 (“Tasterigel” and step cylinder) can be used.



**Fig. 18** An overview of influence factors in XCT. (Adapted from VDI/VDE 2630-1,2, Reiter et al. 2016; Müller 2012; Carmignato et al. 2018)



**Fig. 19** Calibration bodies used for XCT: 13 ruby spheres on carbon fiber-reinforced shafts are fixed on an Al body (a) and ball bar calibration body with 2 ruby spheres mounted on carbon fiber-reinforced shafts (b)

- Scanning parameters: more than 40 parameters are relevant.
- Reconstruction: the reconstruction algorithm and the respective parameters are relevant.
- Reduction or correction of artifacts.
- XCT data processing, in particular preprocessing and proper segmentation/surface extraction.
- Extraction of measurement features from the processed XCT data.

If the whole process chain is aligned in an optimal way the measurement uncertainty for XCT can be determined by using the “Guide to the Expression of Uncertainty in Measurements” (GUM). The GUM is a standardization document that outlines an analytical method to determine and evaluate measurement uncertainty (BIPM 2008). GUM assumes that every measurement can be described by a model. According to recent investigations, the correct formula for the expanded uncertainty in measurements  $U$  is the following (Schmitt and Niggemann 2010):

$$U = k \cdot \sqrt{u_{\text{cal}}^2 + u_p^2 + u_w^2 + b^2} \quad (9)$$

$k$  is the expansion factor (depending on the probability density function of the measurement output quantity).

$k$  is usually chosen to be 2 in many practical cases).

$u_{\text{cal}}$  is the standard calibration uncertainty of the material standard.

- $u_p$  is the standard uncertainty (standard deviation) of repeated measurements.  
 $u_w$  is the standard uncertainty from the production-caused dispersion of workpiece properties (e.g., thermal expansion coefficient, form deviation, roughness, etc.).  
 $b$  is the systematic deviation between the mean measurement value and the calibration value.

The factors  $u_{\text{cal}}$  and  $u_w$  are obtained by using non-statistical data.  $u_p$  and  $b$  can be determined via experiments under repetitive conditions. While  $u_{\text{cal}}$  is dependent on the measurement feature and the calibration process but independent from the parameter setup of the tomographic process,  $u_p$  and  $b$  may strongly vary for different XCT parameter sets and different environmental conditions. The contribution of the measurement process is quantified through the empiric standard deviation  $s_g$  of the feature-related measurement results:

$$u_p = h \cdot s_g \quad (10)$$

This comprises the combined effects of beam generation, attenuation, imaging, reconstruction, segmentation, and geometrical feature extraction. By multiplying the empiric standard deviation by the factor  $h$  it is considered that for a small number of repetitions the measurement uncertainty may be underestimated. This factor ranges from 1 for at least 10 repetitions to 7 for only 2 repetitions.

The influence of the workpiece can have various causes, e.g., thermal expansion due to temperature changes, form deviations, roughness, or changes in the material composition since XCT as a method is based on material-dependent attenuation. The influence of the material composition is difficult to determine by means of experiment because for this purpose various compositions covering the full range of the workpiece are needed. The combined uncertainty for the influence of the workpiece can be calculated as follows:

$$u_w = \sqrt{\Delta L^2 + (b \cdot R_{z,\text{mean}}/2)^2} \quad (11)$$

$\Delta L$  is the elongation of the workpiece and  $R_{z,\text{mean}}$  the mean value of the roughness measured on various spots at the surface of the specimen. The limited voxel size leads to low pass filtering of the surface of the specimen. It is estimated that half of the generated surface lies within the material of the part. Therefore,  $R_{z,\text{mean}}$  is divided by 2 in Eq. 11. If the systematic deviation  $b$  of  $y_{\text{mean}}$  to the calibrated value  $x_{\text{cal}}$  is not compensated, it has to be added to the uncertainty in measurement:

$$b = y_{\text{mean}} - x_{\text{cal}} \quad (12)$$

A comprehensive investigation revealed, for instance, a value of 0.0046 mm for the measurement uncertainty of  $\mu$ -XCT measurements with a voxel size of  $3.6 \mu\text{m}^3$  of a micro gear (Harrer and Kastner 2011). This investigation shows that a measurement uncertainty of one voxel or below can be achieved by XCT.

## Quantitative XCT: Advanced XCT Data Processing and Visualization Methods

XCT currently transforms from a qualitative diagnostic tool to a quantitative characterization method. Quantitative XCT is the combination of XCT with quantitative 3D image analysis (Schladitz 2011; Maire and Withers 2014; Kastner et al. 2017a). Only through preprocessing and data enhancement, segmentation, feature extraction and quantification, rendering of the results, in-depth insights into XCT data a sample may be facilitated. In the beginning of industrial XCT, XCT images were generated mainly for visual inspection. The most important application of quantitative XCT is metrology. Additionally, quantitative XCT is increasingly used for extracting a large variety of characteristics of materials and samples:

- Characterization of pores metallic and polymeric foams
- Porosity evaluation of metals and polymers
- Determination of fiber orientation, diameter and length of fiber-reinforced polymers, as well as their distributions
- Fiber bundle extraction and characterization of technical textiles
- Quantitative data concerning the 3D structure of inhomogeneous metals or other materials (e.g., interconnectivity, sphericity, etc.)
- 3D characterization of isolated discontinuities such as cracks, voids, inclusions, delamination, etc.
- Phase identification and characterization
- Physical and mechanical properties (physical density, crack growth, wear) and, to a certain extent, chemical composition (alloy and phase identification, impurities)

## XCT Data Processing, Analysis, and Visualization

A typical XCT evaluation workflow consists of several steps. A generalization of the XCT data evaluation workflow integrates the following three steps aside from the plain data generation; (1) preprocessing and data enhancement, (2) segmentation, feature extraction and quantification, and (3) rendering of the results (Heinzl et al. 2018).

**Preprocessing and data enhancement:** After reconstruction, pre-processing is often necessary to reduce artifacts and noise in order to increase contrast between different features. For this purpose, artifact correction methods may be applied as well as prefiltering/denoising techniques. As artifact reduction methods have been described earlier in this chapter, we focus here on smoothing and denoising methods. Generally, denoising targets to remove noise and retain original signal in the data regardless of the frequency content of the signal. A group of methods which allow for enhancing noisy XCT images may be found in multi-resolution image denoising. These methods are based on the precondition, that for the correct application of the respective denoising filters, an exact characterization of the imaging system regarding its underlying noise model is required (e.g., Poisson noise, Gaussian noise, etc.).

Results of multi-resolution image denoising techniques have shown, that if the photon transfer curve of the XCT detector is measured or known, multi-resolution denoising methods using wavelet or platelet filters can improve the quality of XCT images considerably (Russ 2002).

Smoothing may be subdivided into global and local techniques. Approaches of global smoothing techniques target a reduction of additive noise and other irregularities in the data, while local (nonlinear) smoothing tries to adapt the smoothing process in accordance with the regional features present in the data. Corresponding smoothing filters range from simple Gaussian or Median filters to advanced bilateral or edge-preserving smoothing filters. A review of filtering methods applicable for preprocessing of X-ray CT images can be found in the work of Heinzl et al. (2018). Extensive smoothing and denoising may reduce segmentation quality by removing true features instead of noise or by generating additional artifacts, and therefore should be applied with care.

**Segmentation, feature extraction, and quantification:** The second step in the generalized workflow for XCT data quantification and visualization is found in segmentation, extraction, and quantification of features in the data. Segmentation partitions an image or a volume into two or more regions and is therefore of core importance for analyzing XCT images. The segmented result is the basis for any further quantification of the features of interest. Here, features of interest may be voids, pores, inclusions, or fibers of which materials scientists require precise characteristics of the individual features such as lengths, diameters, shape factors, etc.

- **Threshold-based segmentation techniques** are most commonly used to distinguish individual regions in a dataset, if these regions feature different gray values. Techniques of this category select one or more suitable values, i.e., the threshold (s), to separate those regions. The thresholds may be empirically determined or computed by an algorithm. For the latter, respective methods are based on the analysis of the histogram shape, correlation between background and foreground pixel/voxel entropies, the similarity of attributes between the gray-scale and binarized image, or higher-order probability distributions and spatial correlations between image pixels. As an example of this category, Otsu's method, tries to maximize the interclass variance while minimizing the intraclass variance in bimodal histograms. While for global thresholding methods the global threshold value is applied on the complete image, local thresholds of respective methods only apply to the specific region they have been computed for. These methods thus utilize a variable thresholding, which may be adapted to the features present in the data, based on local spatial information such as a two-point local covariance function of the image (Oh and Lindquist 1999), edge detection (Sheppard et al. 2004), or others. Regarding threshold-based techniques, the reader is referred to a detailed survey on respective techniques by Sezgin and Sankur (2004).
- **Boundary-based methods** show reasonable performance when its regions are distinguishable by their gray value. Compared to threshold-based techniques, boundary-based methods use (relative) differences of gray values within a defined neighborhood to detect the local borders between regions of interest and only

return regions where edges (i.e., material interfaces) are prevalent. As example of boundary-based segmentation techniques edge detection we indicate here Canny's edge detector as its most prominent candidate. For the separation into individual regions based on the extracted boundaries, typically a succeeding step is required.

- **Region growing methods** build upon the assumption that voxels of a region of interest are connected and typically similar. To segment these regions, region growing methods start from one or multiple seed points or seed regions and grow these regions as long as a homogeneity criterion is valid (Adams and Bischof 1994; Zhigeng and Jianfeng 2007). The watershed transform may also be classified as region-based method as it considers the image as a height field: Similarly, as water filling a valley in the topology of a landscape is being exposed to heavy rain, Beucher and Lantuejoul (1979) successfully transferred this principle to the image segmentation domain, growing regions toward the ridges of the height field.

Aside from conventional segmentation techniques returning a single label for a voxel or region, more recently probabilistic or fuzzy segmentation techniques are also gaining importance. For example, random walker as a probabilistic segmentation algorithm considers the image to be segmented as graph, with pixels as the graph nodes. The edge weight between those graph nodes encodes the similarity between neighboring pixels. The algorithm computes the probability of a random walk starting at a considered node for reaching a seed of the respective label, i.e., the respective material of interest.

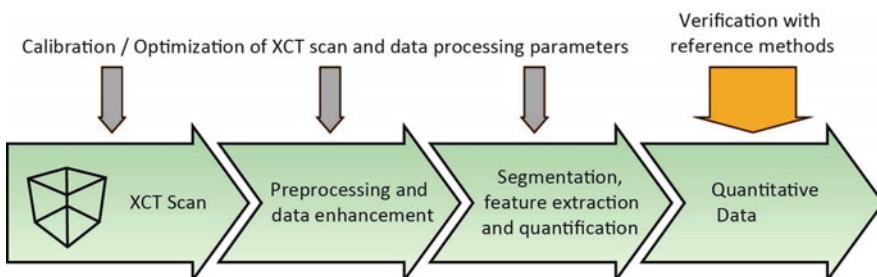
**Rendering of results – visual analysis:** Visualization mapping and rendering of the final results marks the end of the generalized workflow for XCT data processing. The generated (quantitative) data is visualized in an interactive framework, in order to be explored and interpreted by the domain experts (i.e., Visual Analysis). Here it is of core importance to represent and transport the generated information by means of intuitive and easy-to-understand visual metaphors. How the visual analysis system is setup is strongly application specific and thus depends on the underlying data characteristics. A recent state-of-the-art report by Heinzl and Stappen (2017) shows the interrelation of both visual computing and materials science and indicates tasks, techniques, and challenges of the field. In the following paragraphs, we indicate some of these most used visual computing techniques:

- **Volume rendering** is a type of visualization techniques designed to depict 3D volumetric images. Typically, volume rendering algorithms are based on color and opacity transfer functions to map voxels in the data to opacity and color. The main challenge related to volume rendering is that rendering especially of larger volumes (CT images may be up to several hundreds of gigabytes in size) is a computationally costly process. There is a variety of volume rendering techniques which show improved speed or rendering quality.
- **Slicing is a simple** yet highly efficient technique to virtually cut through a volumetric dataset. It thus allows extraction and rendering of arbitrary 2D slices

from a volumetric image. For this reason, slicing is very important and widely used for visual analysis of XCT data, as it provides an intuitive representation of XCT data. In addition, slice images may be compared to different modalities such as microscopy at the same position in the dataset. Similar to volume rendering techniques, slicing techniques are typically implemented in an interactive manner for dataset exploration.

- **Heat maps** are graphical representations of phenomena of interest in a dataset. By using color mapping, the frequency or density of the occurrence of a phenomenon/feature of interest is mapped to colors in the respective space of the heat map. Heat maps are also an important means of data abstraction. For example, features in a 3D dataset may be projected onto a 2D plane. An example is found in porosity maps (Reh et al. 2012), which is an aggregation of all pores in a volumetric dataset in a heat map.
- **Scatter plots** show the correlation of two characteristics or variables in a set of data points. The abscissa and ordinate plot two characteristics of interest from a dataset against each other. The data are rendered as points in the respective scatterplots. Therefore, scatter plots help to manually or automatically identify clusters of features in the data as well as to identify correlations in the data. An extension of single scatter plots is the scatter plot matrix, which aligns multiple plots of different characteristics in a matrix layout in order to explore more than two quantities.

**Quantitative XCT data processing workflow for the extraction of quantitative XCT data:** Quantitative XCT can be described as XCT scan coupled with quantitative 3D image analysis, calibration/optimization with reference parts and/or reference measurements (see Fig. 20). For quantitative XCT, the complete data process workflow from data generation to the final 3D image analysis needs to be optimized and (if possible) calibrated. The XCT measurement and reconstruction parameters as well as data processing methods need to be determined such that the final quantitative values are in accordance with reference measurements ensuring a high repeatability of the whole process chain. Only if there is a proven accordance



**Fig. 20** Data processing workflow for the extraction of quantitative XCT data. Quantitative XCT = XCT + quantitative 3D image analysis + calibration/optimization of the whole process chain. Finally, verification with reference methods is required

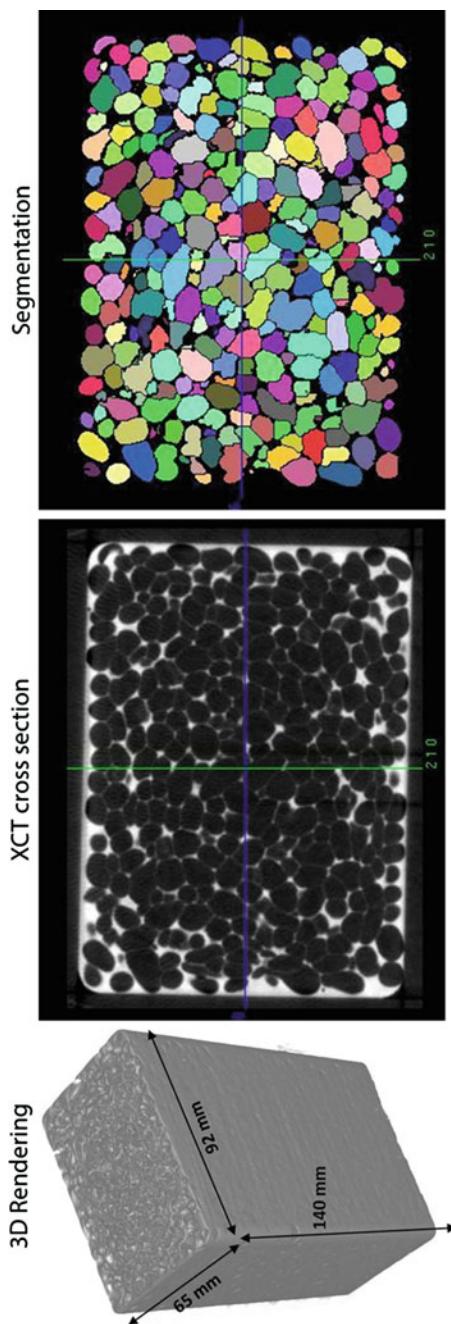
with reference methods XCT may be considered as a quantitative method. If this is not the case, XCT merely provides data for qualitative 3D visualization and characterizations. The reference or verification method depends on the quantitative XCT task: For metrology a coordinate measuring machine may be used as reference; for the characterization of pores in a foam, optical microscopy and laser scanning confocal microscopy are reference techniques of choice; micro-cuts in combination with optical microscopy may be used for porosity measurement in light metallic parts, ultrasonic testing and acid digestion for porosity determination of carbon fiber-reinforced polymers, etc.

## Examples for Quantitative XCT

**Determination of Pore Characteristics in Foams:** XCT imaging in combination with adequate image processing is an accurate and fast method to characterize foams. Figure 21 shows a 3D rendering and the respective XCT cross-sectional image of a closed-cell aluminum foam sample with a cross section of  $65 \times 92 \text{ mm}^2$ . To extract the pore values characteristics of the foam the following data processing pipeline was applied. Gaussian and Laplace filters are used for prefiltering followed by a binarization using the Otsu's method. This binary image supports the watershed algorithm applied on the original XCT gray value image in finding and segmenting the pores. The final image after removing border objects is shown in Fig. 21 right. At the end of this analysis procedure the characteristics for each individual feature as well as for the foam parameters are computed. For each characteristic corresponding histograms and distributions may be computed and plotted and for statistical evaluation. For the foam example this information includes per feature:

- Mean diameter
- Position [x, y, z]
- Volume of each pore
- Surface area
- Shape factor (sphericity)
- Contact surfaces
- Average strut thickness
- Euler numbers and characteristics
- Average and total curvature, etc.

**Determination of porosity of metals:** The detection of heterogeneities such as voids, pores, or cracks in light metals by means of XCT has become a common method of nondestructive testing for scientific and industrial purposes in recent years. Porosity analysis is performed using XCT data in order to assess manufacturing processes. Full information about the internal pore/void shape, position, compactness, size, distribution, and other characteristics may be extracted without destroying the sample, which makes XCT a unique tool for defect evaluation.



**Fig. 21** Quantification of pores in foamed aluminum scanned with a voxel size of  $232 \mu\text{m}^3$ ; left, 3D rendering; center, XCT cross section; and right, segmentation of individual foam cells using watershed segmentation (border objects have been removed); the various colored pores can be distinguished. The borders of the objects are the metaphorical ridges in watershed segmentation and correspond to the aluminum walls between the pores

**Fig. 22** Porosity analysis using XCT data of an Al-die-cast part. Different voids as well as their volumes are represented with different colors according to the color scale ranging from 0 to  $1.8 \text{ mm}^3$ . The voxel size was  $234 \mu\text{m}^3$

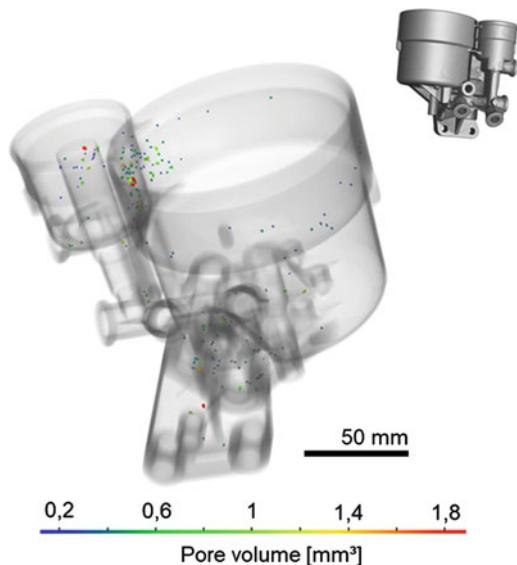
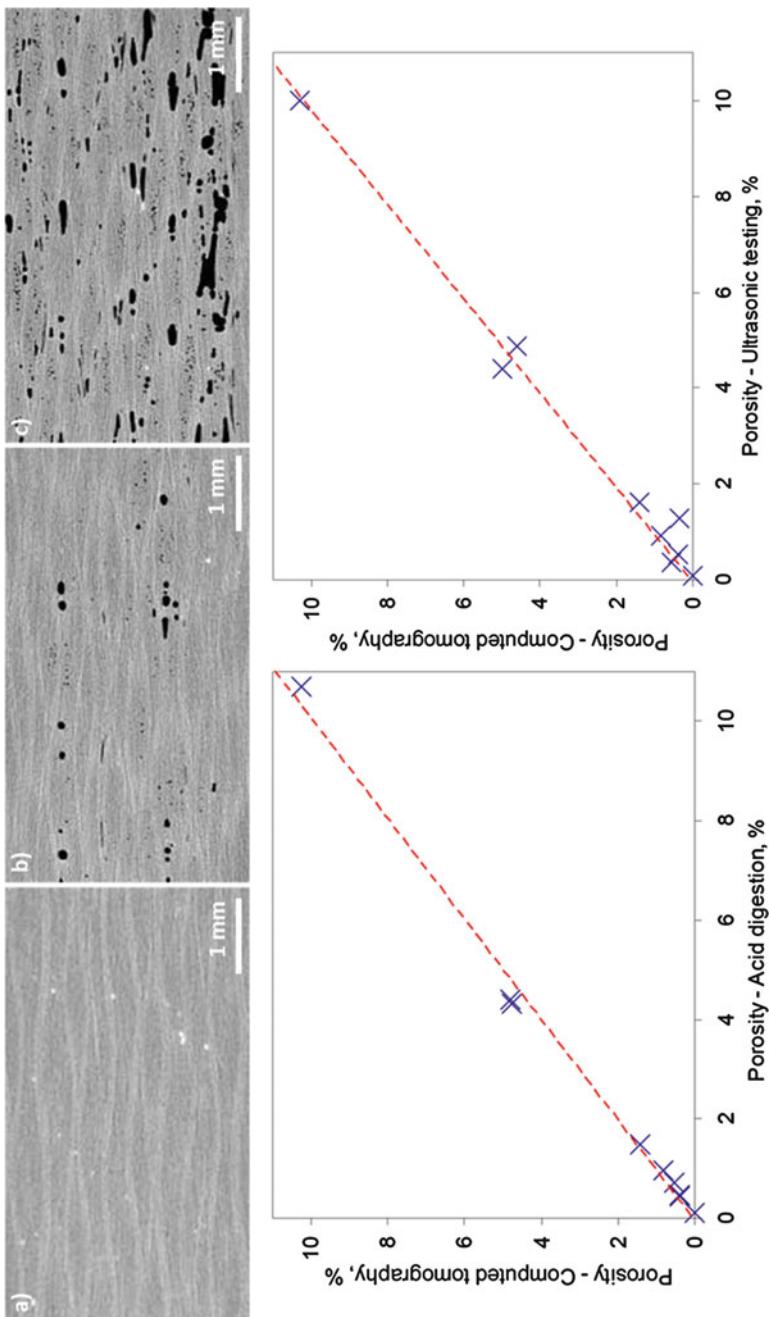


Figure 22 shows porosity analysis of an Al die-casting part in a semi-transparent 3D rendering. The detected shrinkage pores are colored according to the magnitude of the characteristics of interest. In addition, agglomerations become obvious and can be easily identified. For 2D there is a reference method, which is optical microscopy. However, the specimen has to be cut and thus destroyed before it is evaluated.

However, the accuracy of XCT-based porosity measurements and internal defects measurements is still under investigation, and a metrological validation method is needed. Hermanek and Carmignato (2017) recently published a reference object to be used for porosity measurement. It is made of aluminum in a cylindrical body with four cylindrical inserts integrating micro-milled hemispherical features of calibrated sizes resembling artificial flaws. The size range of hemispherical calottes is between 100 and 500  $\mu\text{m}$ . Using this reference object, suitable XCT measurement parameters can be derived, suitable XCT data processing procedures can be developed, and this way the accuracy of porosity measurements by XCT may be evaluated.

**Determination of porosity of fiber-reinforced polymers:** The porosity of fiber-reinforced polymers (CFRP) is a very important factor for practical applications as there is a direct correlation between porosity and mechanical properties, such as shear strength (Kastner et al. 2010b). Typically, the porosity of carbon fiber-reinforced composites for practical use must be lower than 2.5 and, in some cases, lower than 5 vol.%. The standard methods for measuring porosity are currently ultrasonic testing and acid digestion. XCT is a very promising method for accurate determination of porosity and for measuring the spatial distribution of porosities. Examples of cross-sectional XCT data of carbon fiber-reinforced polymer samples with porosity values between 0% and 5% can be found in Fig. 23. For the determination of the quantitative porosity in CFRPs extracted from the XCT data, a calibrated histogram-based segmentation method is used (Kastner et al. 2010b). The average porosity values do not



**Fig. 23** XCT cross-sectional slices of CFRP samples with 0% (a), ~1% (b), and ~5% (c) porosity. Voxel size was  $10 \mu\text{m}^3$ . Comparison of porosity values determined by XCT and acid digestion (left picture) and comparison of porosity values determined by XCT and ultrasonic testing. The correlation between XCT and the reference methods is between 0.98 and 0.99. (Adapted from Kastner et al. 2010)

differ more than 0.3% from the porosity values obtained by the standard methods (ultrasonic testing and acid digestion) as shown in Fig. 23. XCT is not just used to extract the physical value of the porosity. Since XCT provides a 3D representation of the specimen, XCT can be applied to detect and quantify internal defects such as cracks, cavities, or pores. This includes the determination of:

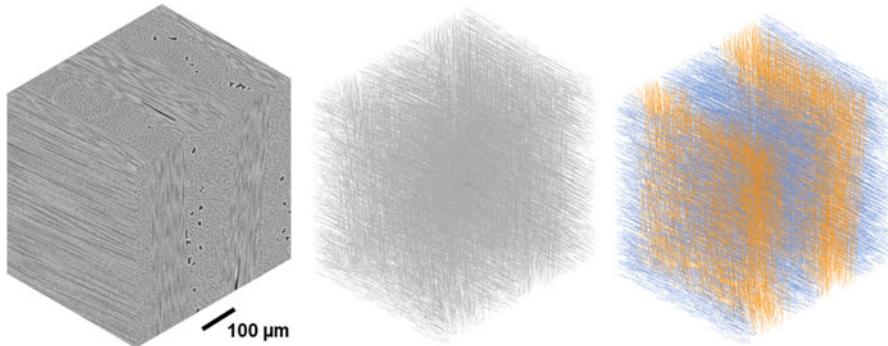
- Defect volume
- Positions of the defects in the sample
- Geometric properties (diameter, volume, sphericity) of the individual defects

Analog to pores, these evaluation possibilities can be applied to quantify properties of inclusions such as foreign particles or fillers in materials.

**Determination of quantitative fiber characteristics for fiber-reinforced polymers:** The contrast between fibers and the polymeric matrix is for most fiber-reinforced material systems (especially for glass fiber-reinforced polymers) high enough to determine the fiber orientation. This is even valid for carbon fiber-reinforced polymers, if the resolution of the XCT measurement is high enough, even though the fibers are typically much thinner (around 6–8  $\mu\text{m}$ ) and the contrast is significantly lower. This opens up a comprehensive characterization of the materials regarding the following properties:

- Fiber length and diameter, as well as the length/diameter ratio (aspect ratio)
- Fiber distribution
- Fiber volume proportion
- Fiber orientation and fiber orientation tensor
- Furthermore, morphological features, e.g., the surface structure of fibers, fiber undulations, or the vascular structure in wood fibers

For the determination of the fiber orientation various methods are used. The method as presented by Salaberger et al. (2011) includes the following consecutive steps: pre-processing, segmentation, medial axis extraction, individual fiber extraction including cluster analysis and final fiber characterization. Other approaches apply Hessian analysis in order to extract the main orientation of the eigenvectors from the data. This way the structure tensor of the fiber orientation may be determined. Determination of the fiber orientation is also possible without segmenting the fibers. Using these concepts, it is possible to exploit the anisotropy of the XCT data to determine the orientations of features in the data, i.e., the orientation of fibers. The example presented in Fig. 24 shows a 3D representation of the XCT data of a carbon fiber-reinforced polymeric (CFRP) sample. The center and the right picture show the extracted fibers. The orientation of the fibers is visualized in blue and orange in the right image in order to distinguish the two main directions of the fibers (perpendicular to each other). We would like to note here that the determination of the fiber orientation can be done much easier as compared to the determination of the fiber length distribution function. Especially at high fiber contents (>30% for glass FRP)



**Fig. 24** XCT data of a carbon fiber-reinforced polymeric sample. Voxel size was  $1 \mu\text{m}^3$ . Left, 3D rendering; center, extracted individual fibers; right, 2 main fiber classes. The orientation of the fibers is visualized by color coding

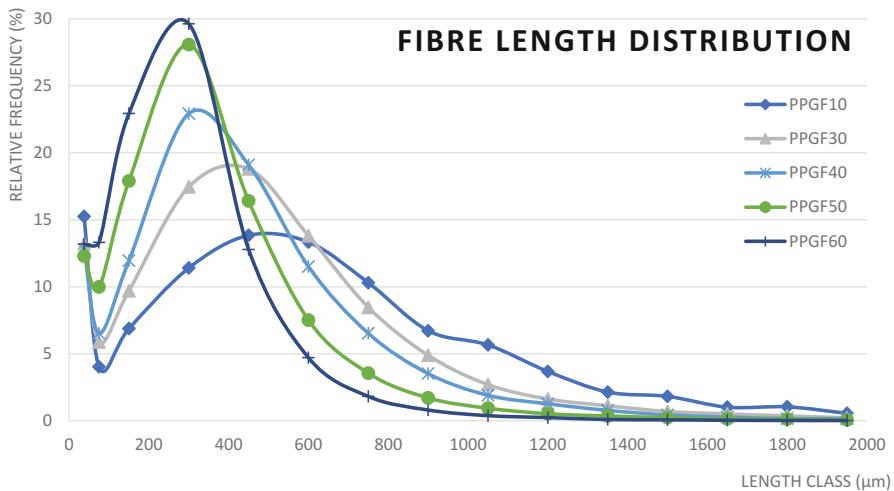
fibers are touching a number of other fibers, which makes a correct segmentation and a correct determination of the fiber's start and end points rather difficult. For the development of respective algorithms XCT simulation is a powerful tool and a suitable ground truth. In addition, verification with standard measurement methods is necessary to ensure the quality of the derived results.

Figure 25 shows fiber length distribution functions of polypropylene granules with 10, 30, 40, 50, and 60 wt.% of glass fiber (PPGF). It is clearly discernible that the fiber length distribution function is shifted to lower values with increasing fiber content. With higher fiber content, more fibers are broken during the manufacturing (compounding) process. Once again, the verification of these quantitative results is crucial. In this case the verification was done by measuring the fiber length distribution function also with a reference method which corresponded with the generated XCT data (Salaberger et al. 2011).

The next example is polypropylene filled with 30% of glass fibers (by weight) (fiber diameter  $12.5 \mu\text{m}$ ) manufactured by injection molding. The injection is in the Z-direction. The XCT data of fiber-reinforced polymers were processed using the pipeline of Salaberger et al. (2011) to determine start and end point of each fiber in the scanned volume, which gives the fiber orientation and the fiber length of each fiber. The extracted fibers are the basis for calculating the three-dimensional orientation tensor. A unit vector  $p$  which lies parallel to the fiber can be used to describe the orientation. The orientation tensor OT describes the orientation of a group of  $n$  fibers and is calculated by the dyadic product of each single fiber:

$$OT = a_{ij} = \frac{1}{n} \sum_{k=1}^n a_{ij}^k = \frac{1}{n} \left( \sum_{k=1}^n p_i^k p_j^k \right) = \begin{pmatrix} a_{xx} & a_{xy} & a_{xz} \\ a_{yx} & a_{yy} & a_{yz} \\ a_{zx} & a_{zy} & a_{zz} \end{pmatrix} \quad (13)$$

The matrix diagonal components  $a_{xx}$ ,  $a_{yy}$ , and  $a_{zz}$  indicate information about the strength of orientations in the corresponding directions. If  $a_{xx} = 1$  then the fibers are oriented parallel to the X axis otherwise if  $a_{xx} = 0$  then the fibers are perpendicular to



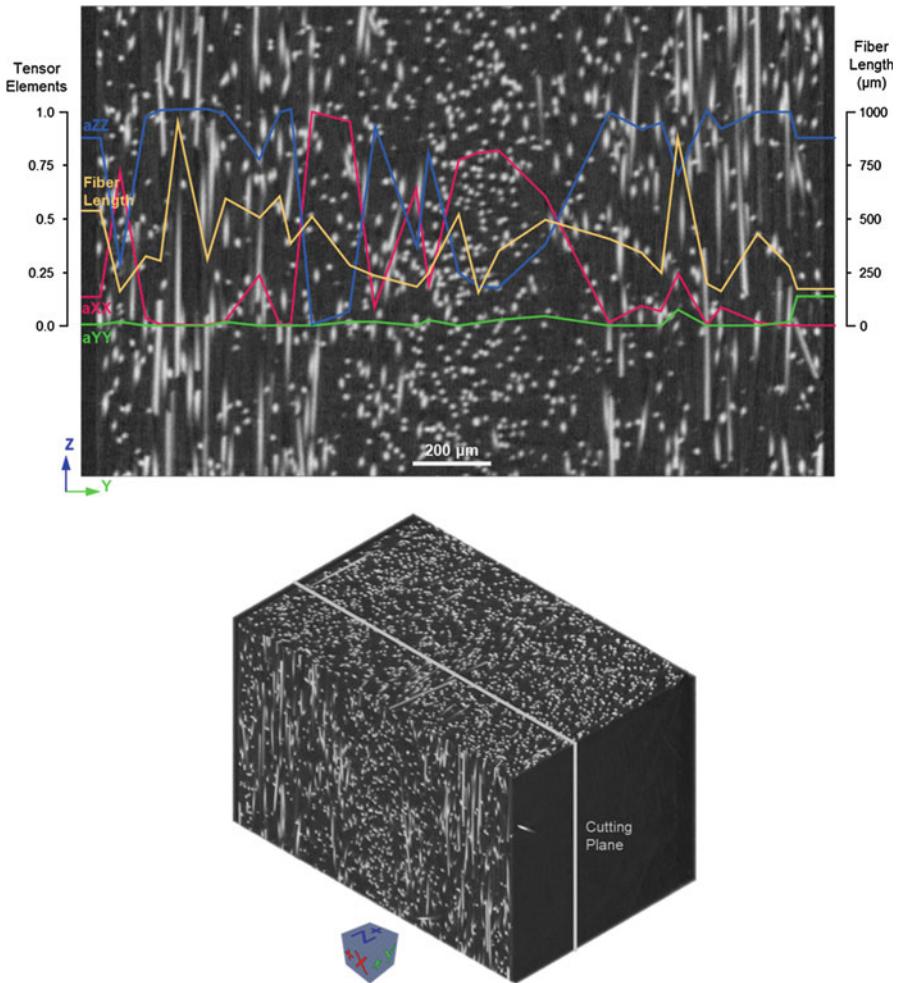
**Fig. 25** Fiber length distribution function of granules with 10, 30, 40, 50, and 60 wt.% of glass fibers (e.g., PPGF10 is polypropylene with 10 wt. % of glass fibers). The fiber length distribution was determined from XCT data with  $2 \mu\text{m}^3$  voxel size

the X axis. Randomly distributed fibers lead to diagonal components equal to 1/3. Moreover, the sum of the matrix diagonal is always 1. The orientation tensor is defined as a symmetrical second rank tensor:

$$\begin{pmatrix} a_{xx} & a_{xy} & a_{xz} \\ a_{yx} & a_{yy} & a_{yz} \\ a_{zx} & a_{zy} & a_{zz} \end{pmatrix} \quad (14)$$

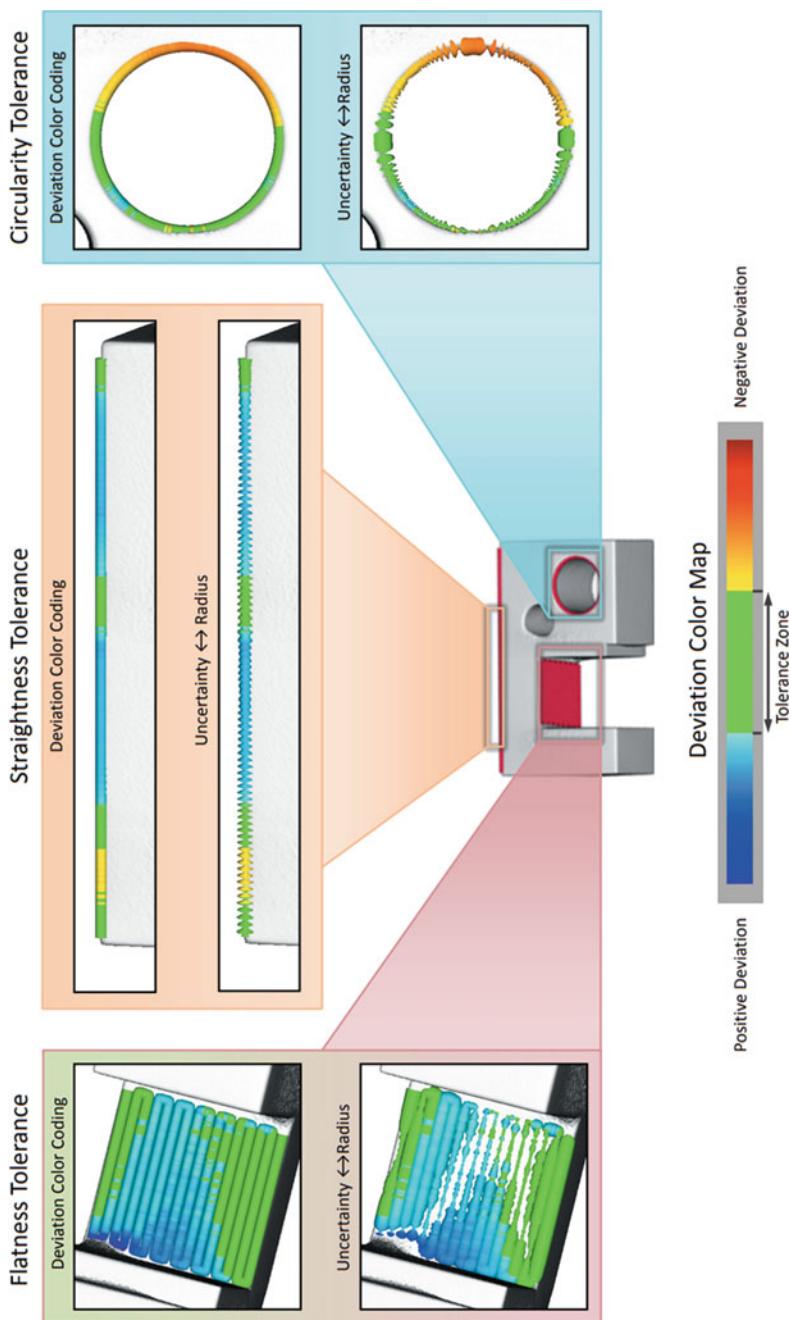
In Fig. 26a an example of XCT cross-sectional pictures of a glass fiber-reinforced polymer sample produced by injection molding is shown together with the mean fiber length (yellow curve) and the orientation tensor elements  $a_{xx}$ ,  $a_{yy}$ , and  $a_{zz}$  (red, green, and blue). The slice image shows a cut through the thickness of an injection molded sheet. The fiber orientation as extracted from the data indicates the expected layered behavior with strong orientation in Z in injection direction near the surface and a transverse orientation in X in the core. This behavior is also illustrated and supported by the orientation tensor elements.  $a_{zz}$  is highest near the surface, which is in the injection direction and  $a_{xx}$  shows high values near the core. The mean fiber length follows this trend, it has minimum in the core regions. Using XCT, quantitative evaluations can be done on slice basis (as shown in the image), on slabs of a certain thickness, on regions of interest (ROI) as well as on the complete specimen.

**Determination of metrological values:** As already described above the second major industrial application of XCT is dimensional measurement. XCT is an alternative to coordinate measuring machines (CMM) which is the standard measurement method in metrology. The applications of XCT for metrology integrate various areas

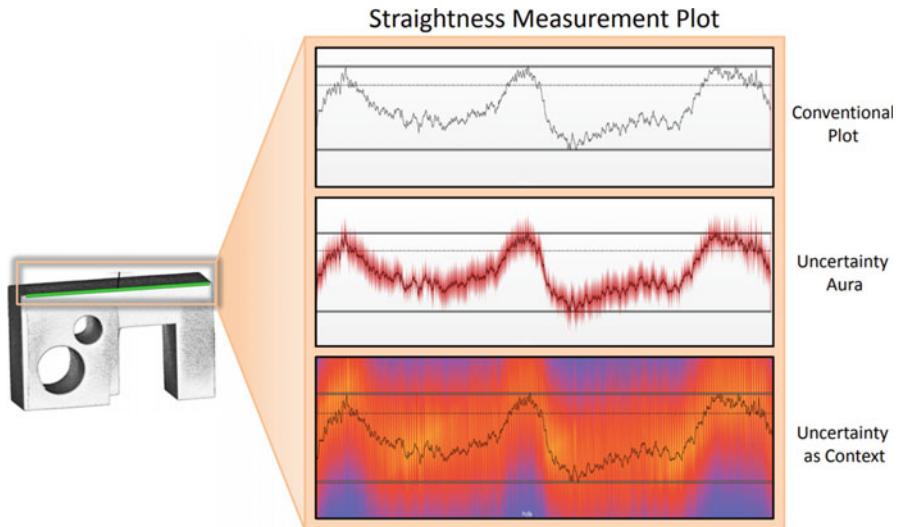


**Fig. 26** Quantitative XCT data illustrating the different regions of an injection molded part with surface layers oriented in Z-direction (orientation of injection molding) and a core layer orientated in X-direction. Voxel size was  $2 \mu\text{m}^3$ . XCT slice together with the mean fiber length and the fiber orientation tensor elements (a) on the respective slice as indicated in the image on the right (b) is shown

such as 3D geometry measurement (especially of hidden structures), measurement of wall thicknesses, actual/nominal comparison, etc. In this short paragraph, we present an example for quantitative XCT and data visualization. Figure 27 shows positive and negative deviations of a flatness tolerance, a straightness tolerance, and a circularity tolerance measured by XCT and a depiction of a straightness tolerance through different visualization methods (Fig. 28) as presented by Amirkhanov et al. (2013) and Amirkhanov (2012).



**Fig. 27** Positive and negative deviations of flatness tolerance, a straightness tolerance, and a circularity tolerance measured by XCT. (Adapted from Amirkhanov et al. 2013)



**Fig. 28** Depiction of straightness tolerance through different visualization methods. (Adapted from Amirkhanov et al. 2013)

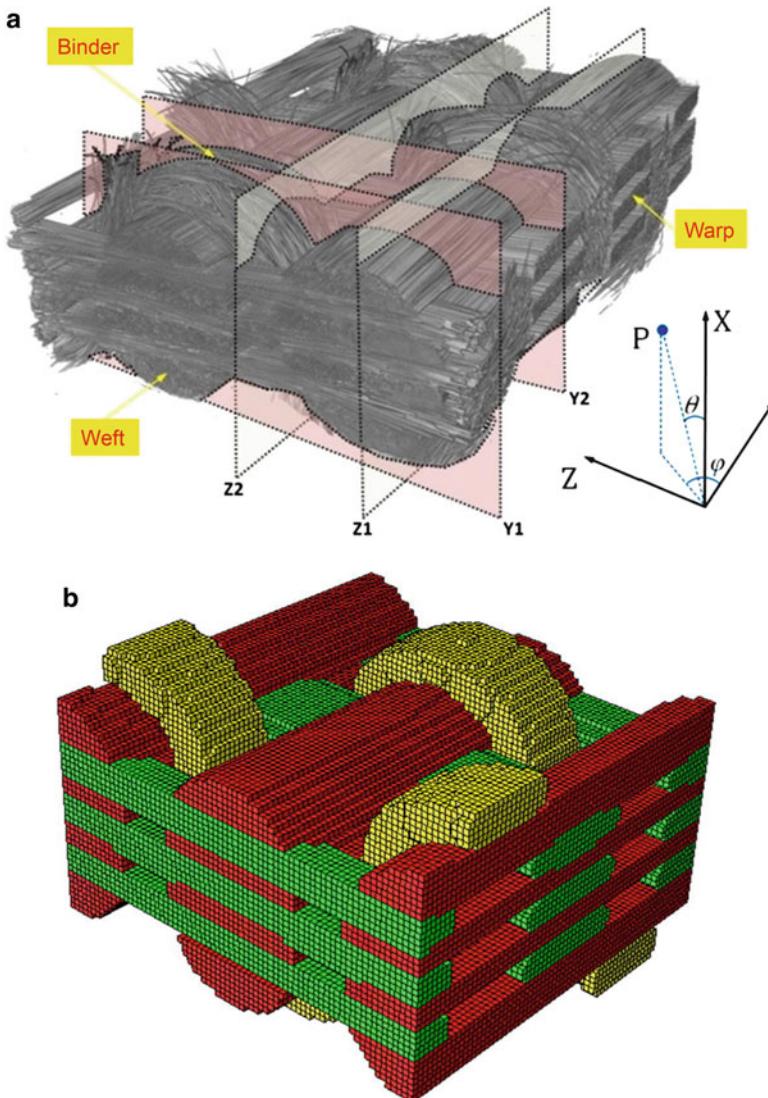
### Quantitative XCT: Combination of XCT and Materials Simulation

The digital nature of tomographic images facilitates the generation of finite element models, which represents the material's microstructure. Voxel-based XCT datasets are therefore converted into finite element meshes, and appropriate boundary conditions and material properties assigned. The digital three-dimensional data obtained from XCT can be converted into finite element meshes in several ways. The most straightforward methods involve converting voxels directly into representative volume elements, trilinear hexahedral elements or in triangle meshes (Oberpeilsteiner et al. 2014; Liu et al. 2017; Garcia et al. 2017). The results can be used to study the material's effective bulk properties, elastic and damage behavior, failure modes and damage progress, permeability of composite structures, the effects of defects/inhomogeneities, and many other properties. This allows a new level of understanding of the intrinsic material behavior.

The first example of XCT materials simulation as presented here is from the field of composites. Textile composites show potential to be applied in a broad range of industries like aerospace, automobile, marine, and civil engineering. A 3D XCT image of a composite material sample provides detailed information about its micro- and meso-structure. By using appropriate image processing methods, XCT data of a composite material sample can be transformed into a three-dimensional finite element model (FEM) (Straumit et al. 2016; Garcia et al. 2017; Liu et al. 2017). This approach to the modeling of composite materials, which is directly based on XCT data, provides a unique opportunity to study mechanical behavior of composite materials taking into account the details of their internal geometry and structure. In this way mechanical properties, damage initiation and development, permeability of the composite

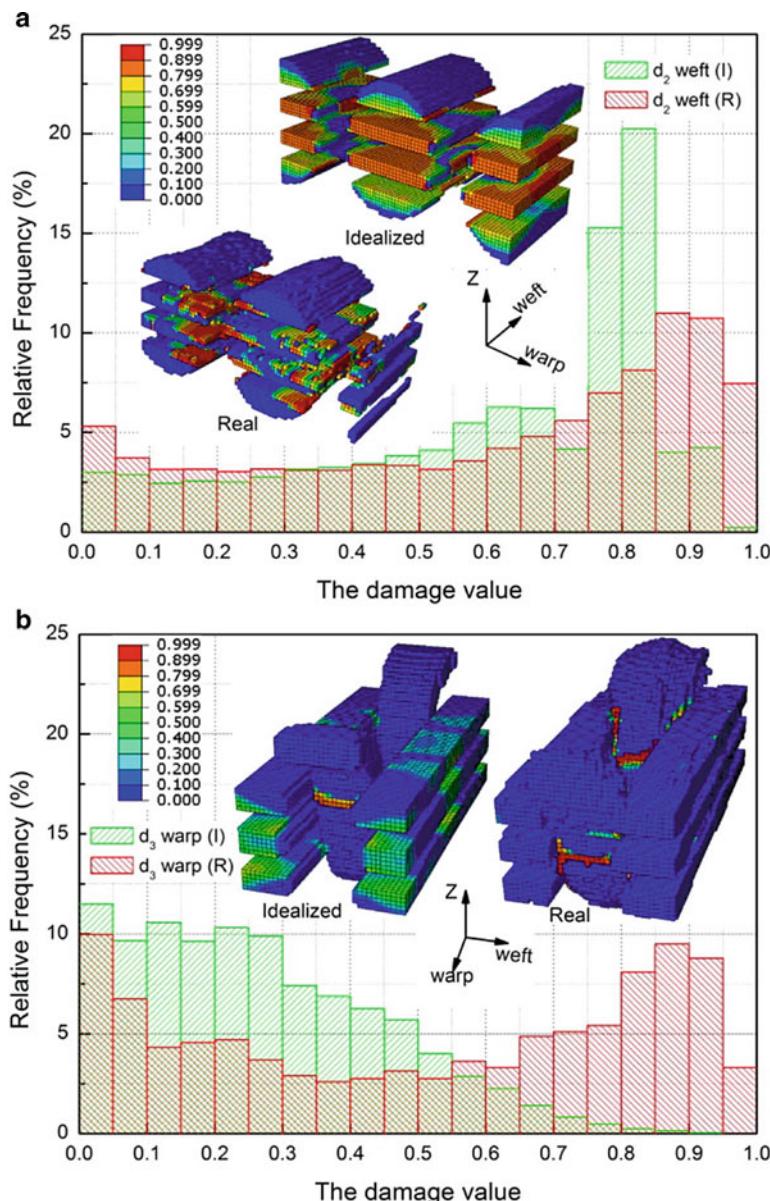
preform, fluid dynamics properties, etc. may be calculated. Figure 30 shows an overview of the workflow to calculate different mechanical and permeability properties of a composite on the basis of a three-dimensional XCT data.

Figure 29 shows the idealized voxel unit cell model of 3D orthogonal weave composite on the basis of XCT data. Warp yarns are oriented along Z-direction while the wefts along are oriented in Y-direction. Furthermore, the damage distribution in weft yarns in the micro-CT based (Real, R) and TexGen (Ideal, I) model is presented just before the failure in Fig. 30. Liu et al. demonstrated that the composite based on the

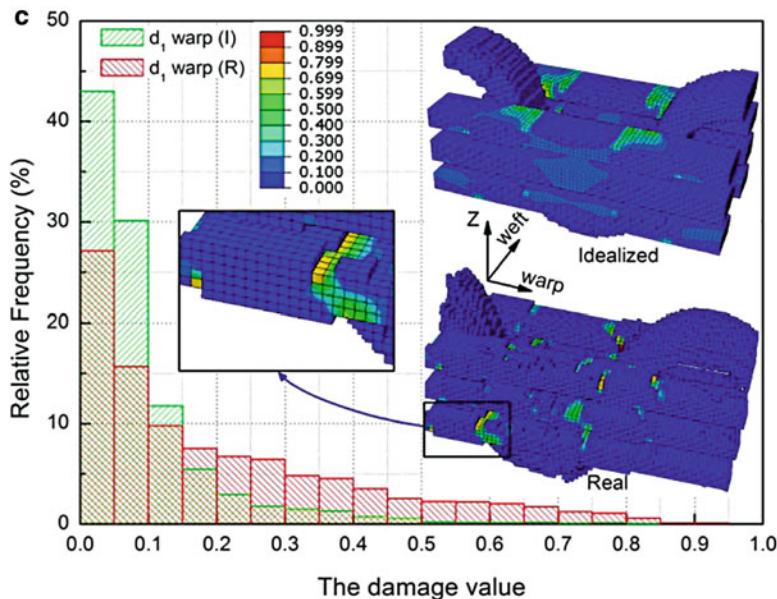


**Fig. 29** Damage modeling of 3D woven glass/epoxy composite (matrix is hidden in the images): (a) micro-CT image; (b) voxel model of the yarns (Liu et al. 2017)

model generated from the XCT images shows better agreement with the experimental results compared to the idealized one (Liu et al. 2017). In comparison with the conventional idealized modeling strategy, the voxel models reproduce the varied yarn cross section and the non-fully-symmetric undulated yarn path much better (Straumit et al. 2016).



**Fig. 30** (continued)

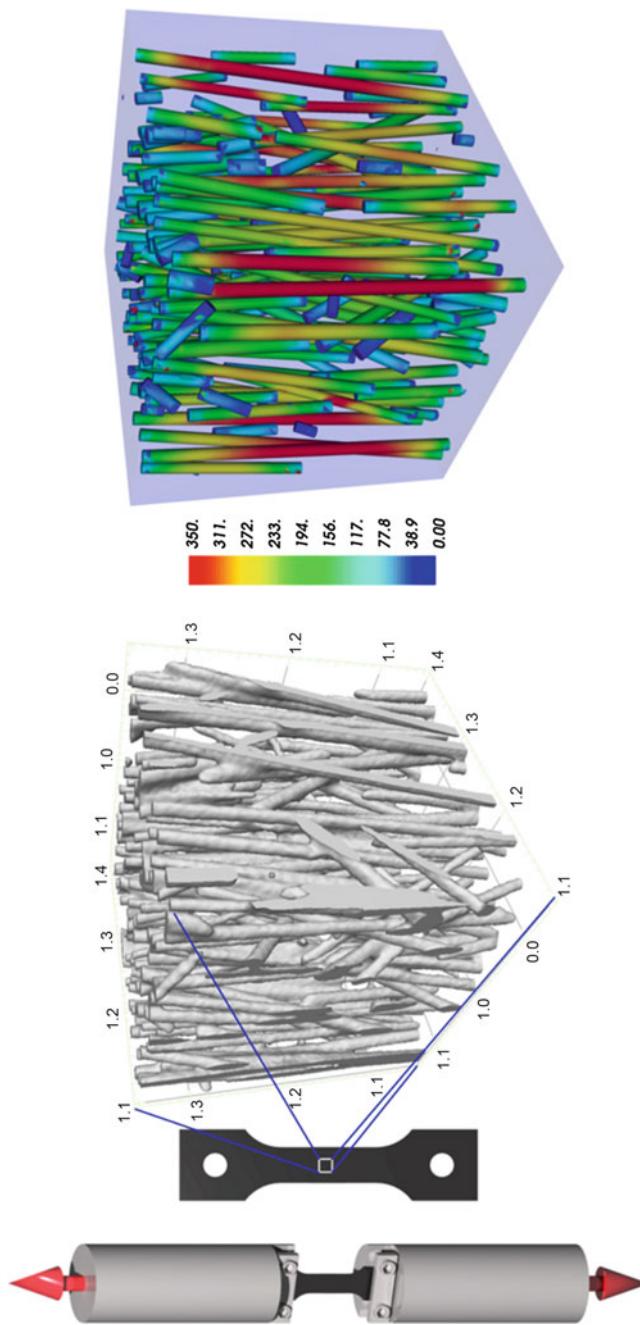


**Fig. 30** Damage distribution in weft yarns in the micro-CT based (Real, R) and TexGen (Ideal, I) model just before the failure (Liu et al. 2017)

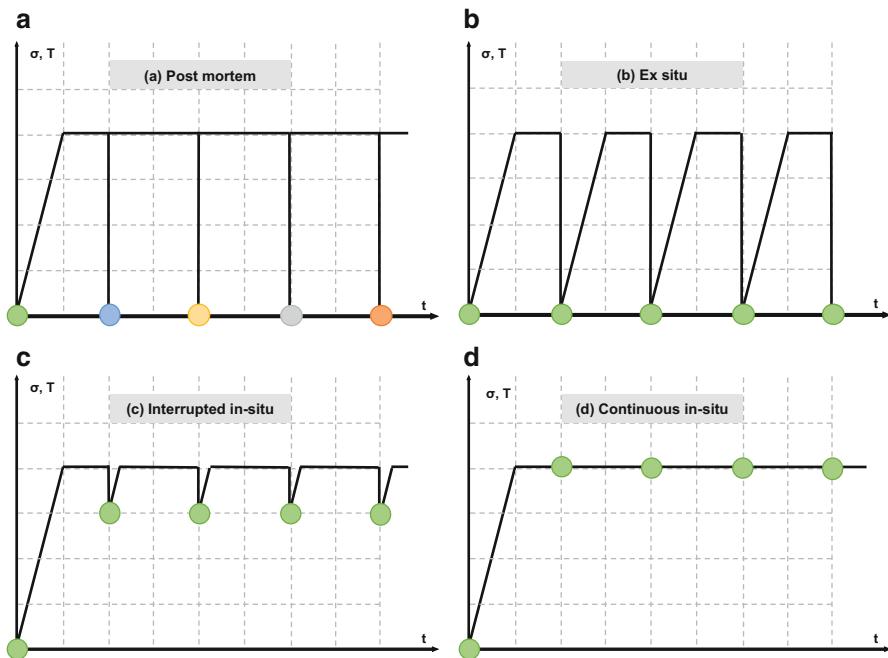
Another example for the use of XCT data in finite element simulation is shown in Fig. 31. In the left picture, an experiment for in situ XCT tensile loading of a GFRP-sample is shown. The fiber extraction pipeline as described above can be used to create an FE model. If load is applied to this virtual specimen one can calculate precisely and visualize forces on the fibers (see Fig. 31 right picture). As demonstrated in Fig. 31 an additional application of XCT for simulation is the evaluation and verification of materials models. XCT can be used, e.g., to measure the size and distribution of inclusions to be compared with materials simulations. XCT can also provide important information about damage initiation and progression (in the form of cracks or local delamination) for the development and verification of material models. A further example of application is the investigation of the effects of defects, where the initial damage state has to be characterized and the load leading to further damage progression has to be identified. This allows the development of a simulation model, which may then be further applied to arbitrary load cases and geometries.

#### 4DXCT/In Situ XCT

Since XCT is nondestructive by its principle, it is predestined for in situ investigations since the process or reaction to be evaluated can be monitored without affecting the sample. The process is observed under real conditions, such as high temperature,



**Fig. 3.1** Schematic representations of a fiber-reinforced tensile sample investigated by in situ XCT (left). The dataset was quantitatively evaluated and used for FEM-simulation of forces (right). The colors represent the strength of the calculated forces (Oberpeilsteiner et al. 2014)



**Fig. 32** (a) Postmortem: different samples at different conditions. (b) Ex situ: same sample at different conditions. (c) Interrupted in-situ: same sample; process is interrupted, (d) Continuous in-situ: same sample under real  $\sigma$ ,  $T$  condition.  $\sigma$  refers to mechanical testing and  $T$  to thermal treatment. Postmortem uses different samples for testing, while all others use the same specimen. (Adapted from Salvo et al. 2010)

in a reaction cell, in a controlled gas atmosphere, under mechanical loads, etc. Typical in situ investigations include tensile and compression testing, cyclic mechanical testing (fatigue), thermo-mechanical behavior of structural materials, phase transitions, physical reactions (e.g., melting, sintering, diffusion), chemical reactions, etc. XCT is one of the most powerful methods for in situ investigations since it generates the complete 3D information, and in most of the cases, the system studied is not affected by the measurement (Salvo et al. 2010; Buffiere et al. 2010; Amirkhanov et al. 2016; Kastner 2016). Since 3D spatiotemporal data are generated, in situ CT is also called 4D X-ray computed tomography (4DXCT). There are several principles of tomography in materials science using 4DXCT as shown in Fig. 32 (Salvo et al. 2010):

- **Postmortem:** several comparable samples are scanned after specific stages of the heat treatment or mechanical testing process. In order to study the influence of the thermal treatment or the mechanical test on the microstructure, several samples must be done.
- **Ex situ:** the procedure is similar as for postmortem tomography but with the difference that the same sample is scanned. This requires sample cooling and

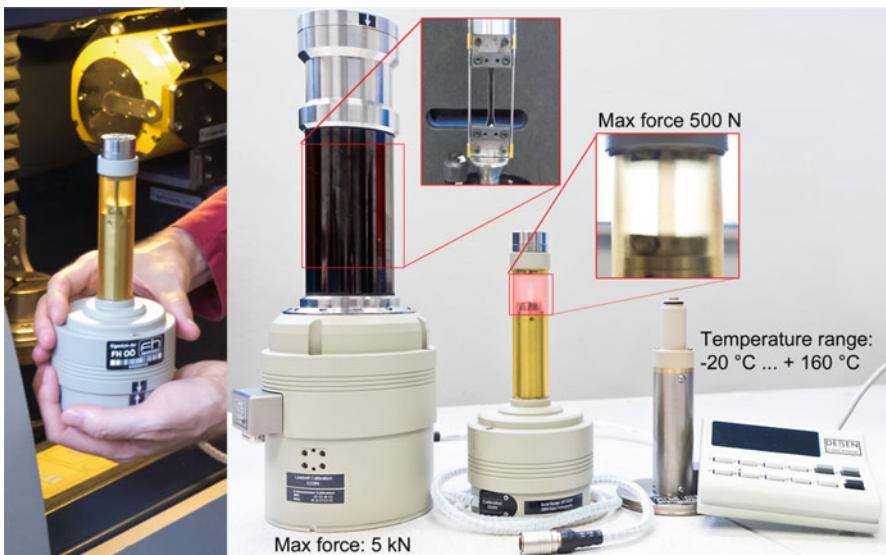
reheating for thermal treatment and stress relaxation for mechanical testing, which may influence the mechanisms to be observed.

- **Interrupted in situ:** the procedure extends on the ex situ tomography principle and interrupts the ongoing process for scanning (temperature or force is kept constant for the scan). This technique requires specific testing devices (furnaces, compression-tension machine) mounted directly on the XCT device. The same sample is scanned, but mechanical tests need to be interrupted during the scan which again may affect the mechanisms.
- **Continuous in situ:** in this procedure, the same sample is scanned without interrupting the mechanical test or the thermal treatment. This requires fast scans ensuring that the microstructure of the material does not change too much during the scan.

Laboratory XCT devices with resolutions close to  $1 \mu\text{m}$  can be used for many in situ investigations. However, due to the typical measurement times of 30–200 min, usually only interrupted in situ investigations (aside from postmortem and ex situ test) are possible. Synchrotron XCT reveals substantial advantages when fast processes or processes at high temperatures have to be investigated. Therefore, continuous in situ XCT is typically performed at synchrotron beamlines. Concerning in situ XCT at different loading levels, the major experimental challenge is the integration of a loading test machine in an XCT device. No standardized test set-ups are established and available yet. Therefore, the experimental feasibility is still a challenge and needs to be validated. In situ techniques are often combined with theoretical calculations and simulations, in order to advance the understanding of material properties and processes. Typical experimental setups for interrupted in situ tensile tests and temperature-dependent experiments at the XCT devices are shown in Fig. 33. These loading stages can apply a maximum force of 500 N with a minimum voxel size of around  $(2 \mu\text{m})^3$  and a maximum force of 5 kN with a minimum voxel size of about  $(4 \mu\text{m})^3$ . The temperature rig operates at temperatures between  $-20^\circ\text{C}$  and  $+160^\circ\text{C}$  with voxel sizes down to  $(2 \mu\text{m})^3$ .

An important application of 4DXCT is the analysis of defects in glass fiber-reinforced polymers during tensile tests. The damage introduced during tensile testing and the damage mechanisms can be clearly detected by XCT analyses. The main mechanisms of this material system are found in fiber pull-outs, fiber fractures, fiber-matrix debonding, and matrix cracks. Figure 34 (top) shows XCT cross sections of these typical damage mechanisms. The resolution in this application is high enough so that the different damage mechanisms can be clearly distinguished and in addition, a quantitative classification is possible (Amirkhanov et al. 2016). In the right image of Fig. 34 (bottom), the number of defects (fiber pull-outs, fiber fractures, matrix fractures, and fiber-matrix debonding) versus the applied force is plotted. The XCT scans were performed in an in situ loading stage exerting up to 90% failure force of the material system and with a voxel size of  $2 \mu\text{m}^3$ . These images show that fiber pull-outs and matrix fractures are the dominant defects for this kind of material, whereas debondings play almost no role.

The next example demonstrates how 4DXCT can be used to understand fatigue crack growth in aluminum alloys (Williams et al. 2013). The mechanisms by which fatigue

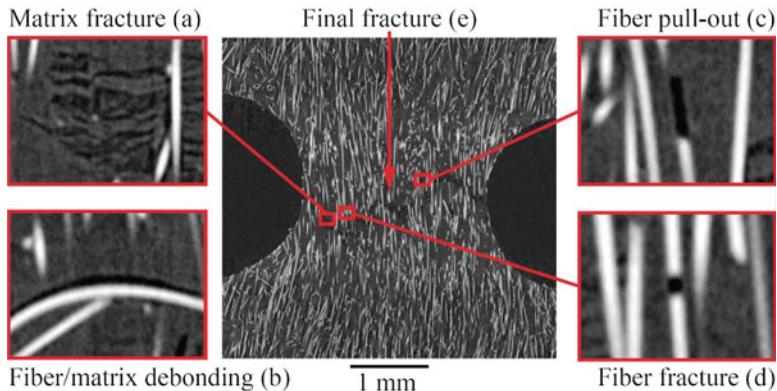


**Fig. 33** GE Nanotom S (180 keV) with 500 N tensile stage from DEBEN (left). Overview of available DEBEN in situ tensile stages (5 kN and 500 N) and cooling stage ( $-20\text{ }^{\circ}\text{C}$  to  $+160\text{ }^{\circ}\text{C}$ ) for XCT applications at FHW

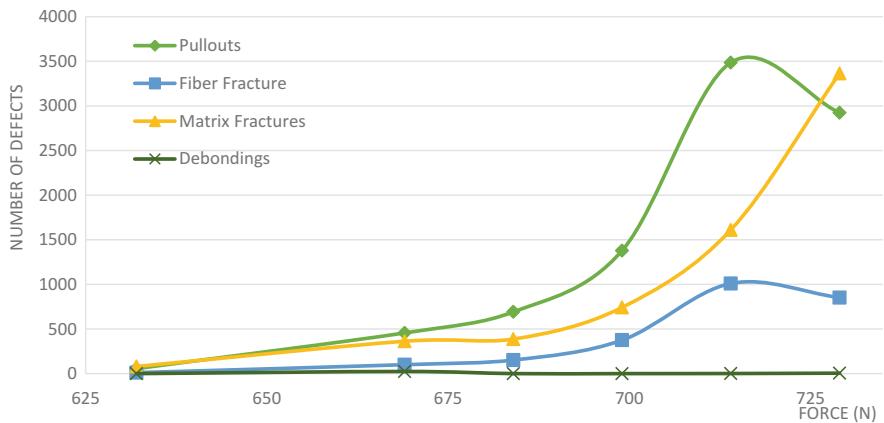
crack growth takes place in metallic materials is still a topic of research. XCT is an ideal technique for extracting the fatigue crack growth process quantitatively in three dimensions and, more importantly, for conducting in situ fatigue crack growth experiments. In situ measurements of crack opening displacements were obtained for the 7075-T6 aluminum alloy (Al6Zn3Mg2Cu), illustrating the possibilities for also quantifying fatigue crack closure. Figures 35 and 36 show a 3D rendering of a 7075-T6 aluminum specimen and the fatigue crack growth profiles as a function of cycles and the 3D reconstructions of fatigue crack through brittle inclusions. The scans were performed at the beamline 2-BM at the Advanced Photon Source at Argonne National Laboratory. An X-ray energy of 24 keV and a voxel size of  $1.8\text{ }\mu\text{m}^3$  were used. Quantitative microstructural analysis enabled an assessment of the role of brittle inclusions on fatigue crack propagation since the contrast between the Al-alloy and the inclusions was high enough for segmentation. It was observed that the inclusions lead to a significant increase in preferential crack growth.

## Phase-Contrast XCT

An emerging XCT method which provides more insights than conventional attenuation-based tomography is called phase-contrast XCT. Phase-contrast XCT can mainly be divided into at least five different methods: propagation-based phase-contrast XCT, Talbot-Lau grating interferometer XCT, crystal interferometer XCT,



**FORCE (N) VS NUMBER OF DEFECTS**

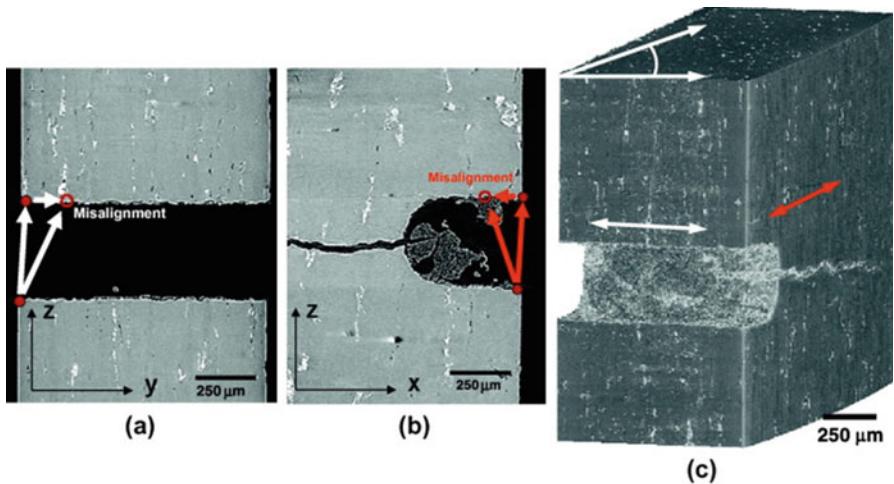


**Fig. 34** Cross-sectional XCT images of typical damage mechanisms in glass fiber-reinforced polymers: matrix fractures, fiber fractures, fiber/matrix debonding, and fiber pull-outs (top). Defects color-coded according to their type: fiber fractures red, matrix fractures blue, fiber pullouts yellow (right image). The XCT measurements were performed using an *in situ* loading stage up to 90% failure force and with a voxel size of  $2 \mu\text{m}^3$ . The number of defects is plotted against the applied force (bottom)

edge illumination, and analyzer-based imaging (Pfeiffer et al. 2006; Zhou and Brahme 2008; Yashiro et al. 2010; Kastner et al. 2012, 2017; Malecki 2013). We will concentrate on the first two methods in this work as they are the most relevant ones for practical applications.

### Propagation-Based Phase-Contrast XCT

The physical principle of the image formation in conventional industrial and medical X-ray computed tomography is based on the absorption of X-rays. However, the



**Fig. 35** Al6Zn3Mg2Cu specimen with an initial crack to study fatigue crack growth. 2D slices from X-ray tomography data showing types of misalignment: (a) crack growing into the plane, (b) side surface of the specimen, and (c) 3D rendering showing axial and angular misalignment (Williams et al. 2013)

thickness or density variation of the specimens do not only lead to changes in X-ray intensity, but also to changes of the phase of X-rays passing through the object. This behavior can be described in Eq. (15) by the complex X-ray refractive index  $n$ :

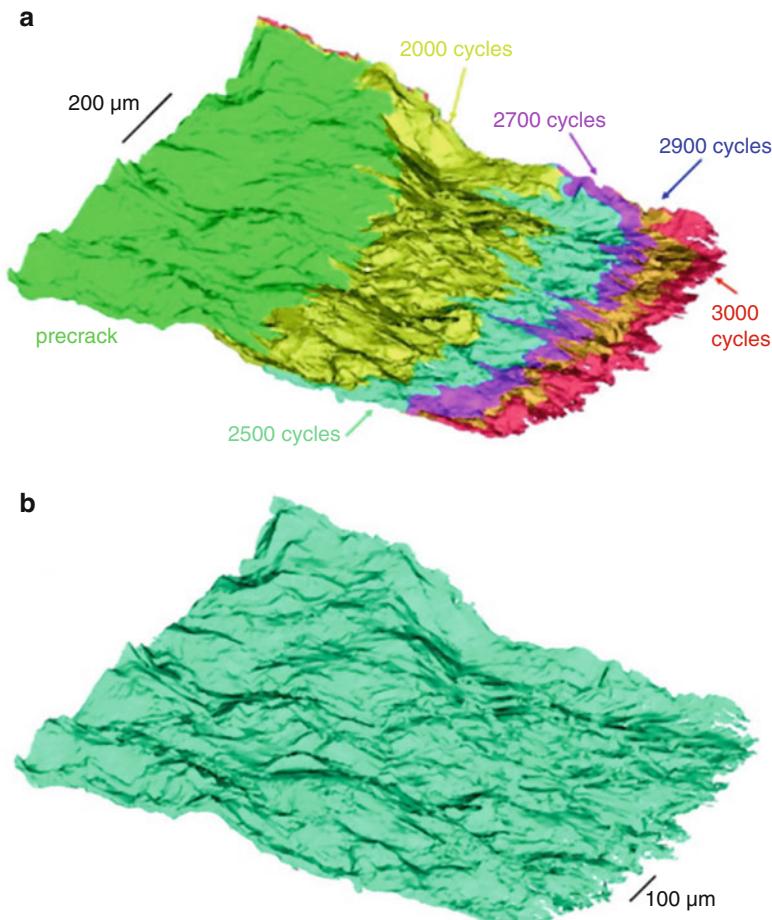
$$n = 1 - \delta + i\beta \quad (15)$$

$\delta$  is the refraction index and  $\beta$  the absorption index. The phase  $\varphi$  of an X-ray beam is changed on passing through a material along the X-ray path in accordance with the following equation:

$$\varphi = -\frac{2\pi}{\lambda} \int \delta(z) dz \quad (16)$$

$\lambda$  describes the mean wavelength of the X-ray spectrum and  $z$  the sample thickness. The integral is computed over the X-ray path through the specimen. Both the refractive and the absorption indices depend on the energy of the X-rays and on the material composition of the specimen. Since the  $\delta/\beta$  ratio for low-density materials is usually higher than that for high-density materials, the phase contrast is most pronounced for materials with a low density around or below  $1 \text{ g/cm}^3$ . The change of phase in the object also leads to a change of the direction of X-rays due to refraction. This results in an upward and downward overshooting of the gray values across a certain edge, which is typical for phase-contrast effects.

When polychromatic laboratory X-ray sources are used, phase information is typically unavailable. This is due to the blur induced by the finite X-ray source size (as well



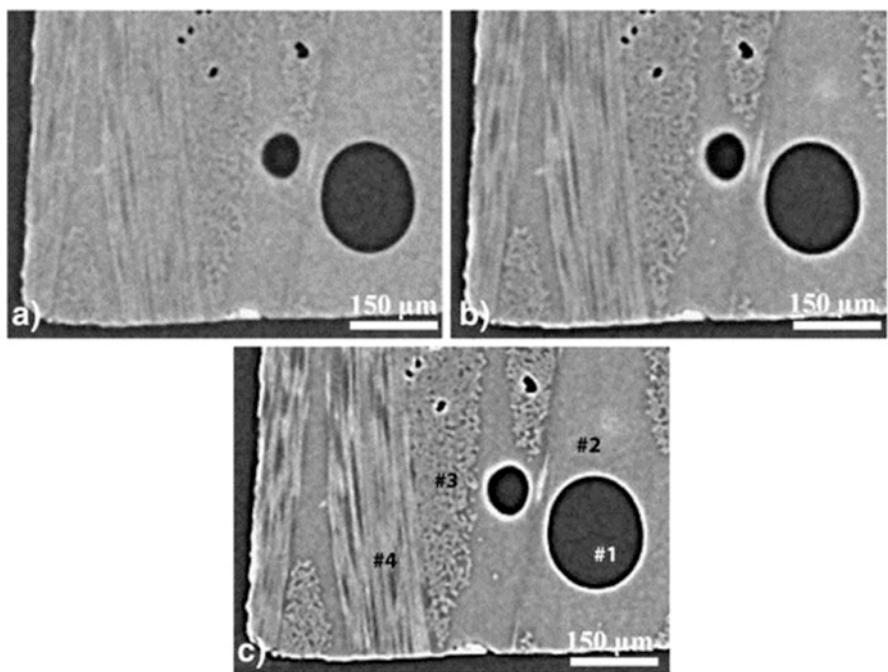
**Fig. 36** 3D fatigue crack growth profiles: (a) crack profiles as a function of cycles and (b) full crack growth profile at 3000 cycles (Williams et al. 2013)

as the limited spatial coherence) and by the polychromatic spectrum of the beam. In special cases even for cone beam geometry using a polychromatic source, phase-contrast effects may occur under special conditions. The spatial coherence length  $L_{coh}$  can be represented according to Zhou and Brahme (2008) by the following formula:

$$L_{coh} = \frac{\bar{\lambda} \cdot \text{FOD}}{s} \quad (17)$$

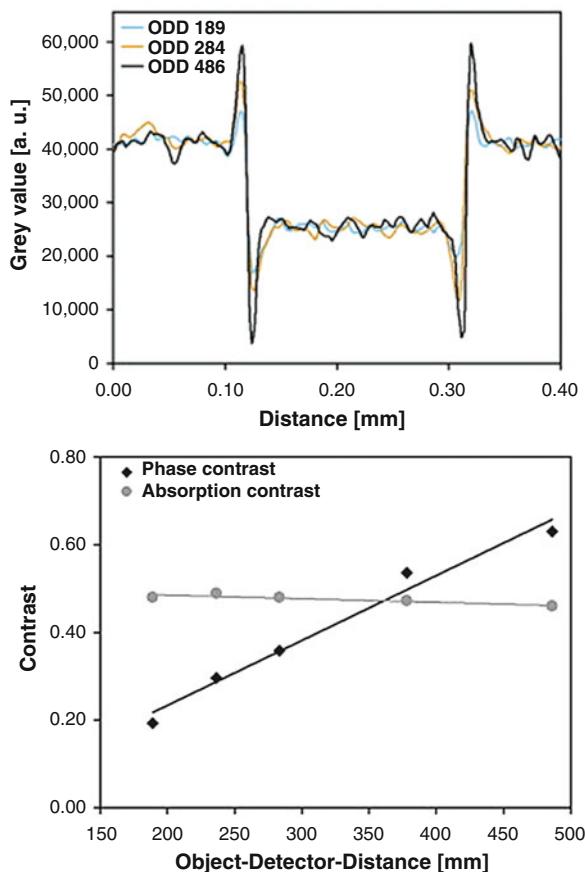
FOD here denotes the distance between the focal spot of the X-ray source and the object, and  $s$  is the focal spot size. Equation (17) shows that maximum phase contrast can be obtained. When the wavelength is large, the focal spot size is small and the distance between the source and the object is large.

In practice a conventional cone beam XCT system with a polychromatic X-ray source without any X-ray optics or collimator can show phase-contrast effects when the focal spot size of the tube is small enough and the distance between the object and the detector is high enough (Kastner et al. 2012): If the focal spot size is around 1  $\mu\text{m}$  or below, the spatial coherence length is high enough so that edge effects, i.e., phase-contrast effects, occur. The influence of propagation-based phase contrast on the resulting XCT data and on the detectability of details is shown in Fig. 37. Here the distance between the object and the detector is increased from 189 to 486 mm, which results in an increase of the edge effects and detail detectability. The left picture in Fig. 38 shows XCT gray value profiles along a big pore of the CFRP sample measured with ODD = 186 mm and ODD = 486 mm, and in the right picture, the phase contrast versus different object-detector distances is plotted. It is clearly visible that the edge effects increase with increasing ODD. The propagation-based phase-contrast effect can be used to increase the detectability of low-absorbing materials or of material transition with a small change in the absorption coefficient.



**Fig. 37** Cross-sectional XCT data of a carbon fiber-reinforced polymer sample at different object-detector distances (ODD) and a constant voxel size of  $2.75 \mu\text{m}^3$ : (a) ODD = 189 mm, (b) 283.5 mm, and (c) 486.25 mm. The phases are pores (#1), polymer resin (#2), carbon fibers perpendicularly to the section plane (#3), and carbon fibers parallel to the section plane (#4) (Kastner et al. 2012)

**Fig. 38** Gray value profiles along a big pore of the CFRP sample measured with different object-detector distances and a constant voxel size of  $2.75 \mu\text{m}^3$ . (b) Phase contrast versus object-detector distances evaluated from gray value profiles from (a) (Kastner et al. 2012)



### Talbot-Lau Interferometer Phase-Contrast XCT

One of the most important and most recent innovations in X-ray technology was the introduction of the Talbot-Lau grating interferometry (TLGI). Fifteen years ago, the Talbot-Lau effect was first utilized for X-ray imaging by utilizing monochromatic synchrotron radiation. The additional use of a source grating facilitated the introduction of TLGI to polychromatic X-ray sources in 2006 (Pfeiffer et al. 2006; Yashiro et al. 2010; Prade et al. 2017). A Talbot-Lau interferometer is formed by three transmission gratings combined with a state-of-the-art X-ray apparatus with low coherence. Three complementary characteristics of the specimen are provided in a single scan using TLGI:

1. Attenuation contrast (AC) due to absorption
2. Differential phase contrast (DPC) due to refraction
3. Dark-field contrast (DFC) due to scattering

AC is equivalent to conventional X-ray imaging, it provides information about the attenuation of the X-ray beam intensity through the specimen. DPC is related to the index of refraction and image contrast and based on the local deflection of the X-ray beam. Finally, DFC reflects the total amount of radiation, which is scattered at small angles by microscopic inhomogeneities in the sample (e.g., by particles, pores, and fibers). Depending on this microstructure, the scattering signal shows a preferred direction which is perpendicular to the local orientation and reflected by the measured dark-field signal.

**Theoretical background of Talbot-Lau grating interferometer XCT:** Conventional X-ray imaging uses the drop-in intensity through attenuation. This drop-in intensity is caused by an object in the X-ray beam. The radiation is treated in this case as rays similarly to in geometrical optics. When X-rays pass through a specimen, not only their amplitude but also their phase is modified. So instead of treating the radiation as simple rays, X-rays can also be considered as electromagnetic waves. As described in Eq. (15) an object can be described by its complex refractive index  $n$  (= real part of the refractive index  $\delta$  and the imaginary part  $\beta$ ). In contrast to visible light the real part of the refractive index  $n$  for different media is generally of the order of  $10^{-5}$ . Therefore, the refraction angles at the boundaries between two isotropic media computed with Snell's formula are also very small. Thus, refraction angles of X-rays passing through an object cannot be detected directly. They are usually determined indirectly through interference patterns between diffracted and undiffracted waves which are generated by spatial variations of the real part of the refractive index. The impact of the refraction index on the behavior of the wave may be demonstrated using a wave propagating in an arbitrary medium with fixed refractive index  $n$ . For reason of simplicity, a monochromatic planar wave with the wavelength  $\lambda$  is assumed. The wave then propagates in a direction normal to the surface of the medium, named  $z$  in this example. The scalar wave function is then:

$$\Psi(z) = E_0 \cdot e^{inkz} = E_0 \cdot e^{i(1-\delta)kz} e^{-\beta kz} \quad (18)$$

$\delta kz$  determines the phase shift and  $e^{-\beta kz}$  is an exponential decay factor decreasing the amplitude  $E_0$  of the wave. In more general terms, the total phase shift  $\varphi$  of the beam propagating a distance  $z$  may be computed by integrating over the propagation length. The phase shift is equal to the projection of the decrement of the real part of the refractive index in imaging direction. This fulfills the requirement of the tomographic principle, that the input data to a reconstruction algorithm should be a projection of a quantity  $f$  that conveys the structural information inside a sample. This means that a map of the real part of the refraction index  $\delta(x,y,z)$  can be reconstructed with standard techniques such as filtered backprojection in phase-contrast imaging. This principle is analog to conventional X-ray computed tomography, where also a map of the imaginary part of the refraction index can be retrieved. Paganin et al. (2002) introduced a method for simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object which is robust, non-iterative, and computationally efficient. As the part of the refractive index  $\delta$  and the imaginary part  $\beta$  have different origins, they are different for most

**Table 5** Dependence of refractive index  $\delta$  and the imaginary part  $\beta$  on atomic number  $Z$  and Energy  $E$ . (Ref. [https://en.wikipedia.org/wiki/Phase-contrast\\_X-ray\\_imaging](https://en.wikipedia.org/wiki/Phase-contrast_X-ray_imaging))

	Atomic number Z of the material	Energy of the X-rays E
Refractive index $\delta$ proportional to	$Z$	$1/E^2$
Imaginary part $\beta$ /Attenuation proportional to	$Z^4$	$1/E^4$

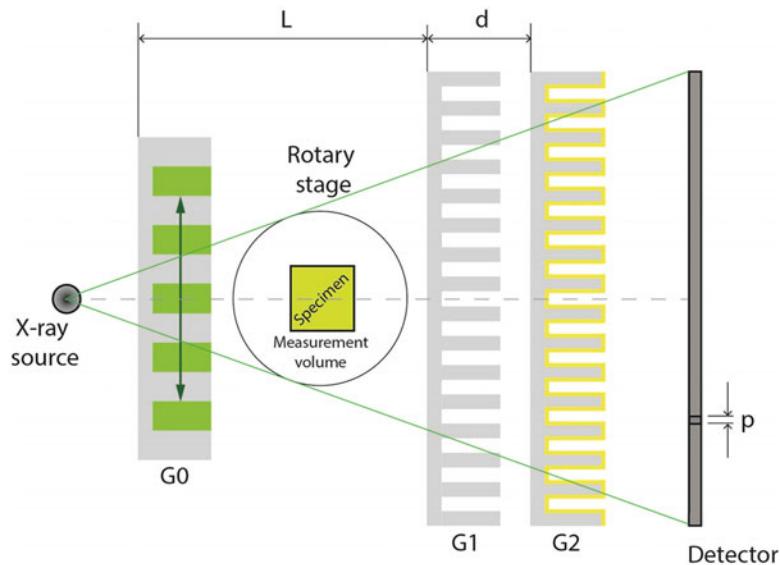
materials and spectral ranges. The ratio between the refractive index  $\delta$  and the imaginary part  $\beta$  can be as high as 100–2000 in the energy range between 10 and 100 keV (Malecki 2013). This means that the phase shift of an X-ray beam propagating through media may be much larger than the loss in intensity. This is also the reason that phase-contrast imaging is more sensitive to density variations in the material than absorption imaging. Table 5 summarizes the dependencies of the refractive index  $\delta$  and the imaginary part  $\beta$  on the material value atomic number  $Z$  and the energy  $E$  of the X-rays.

The dark field signal has its origin in small angle X-ray scattering. It can be described by the directional X-ray scattering coefficient  $D(x,y)$ :

$$D(x,y) = e^{-\int \mu_d \cdot (x,y,z) dz} \quad (19)$$

Here,  $\mu_d$  denotes the scattering coefficient. The dark field signal also depends on the size of the scattering structures. The highest DFC intensity is achieved when the structure size is similar to periodicity of the gratings. Due to the resemblance between the linear absorption coefficient  $\mu$  and the linear scattering coefficient, the tomographic reconstruction of the dark-field signal is done analogue to the reconstruction of the standard absorption. Because of the similarity to Lambert-Beer's law, it is possible to reconstruct a scattering tomogram with conventional signal reconstruction methods by simply exchanging the transmission data with the dark-field data (Jensen et al. 2010; Malecki 2013). However, the dark-field signal strongly depends on the relative orientation between the sample's sub-pixel structures and the grating lines of the interferometer. The dependence on the relative orientation contradicts the assumption of an isotropic scattering contribution produced by the sample when rotating it during a dark-field tomography scan. Only isotropic structures would cause constant scattering. Thus, the simple tomographic approach for dark-field imaging is limited to isotropically scattering samples only. If a sample is not scattering isotropically artifacts might occur in the DFC-XCT data.

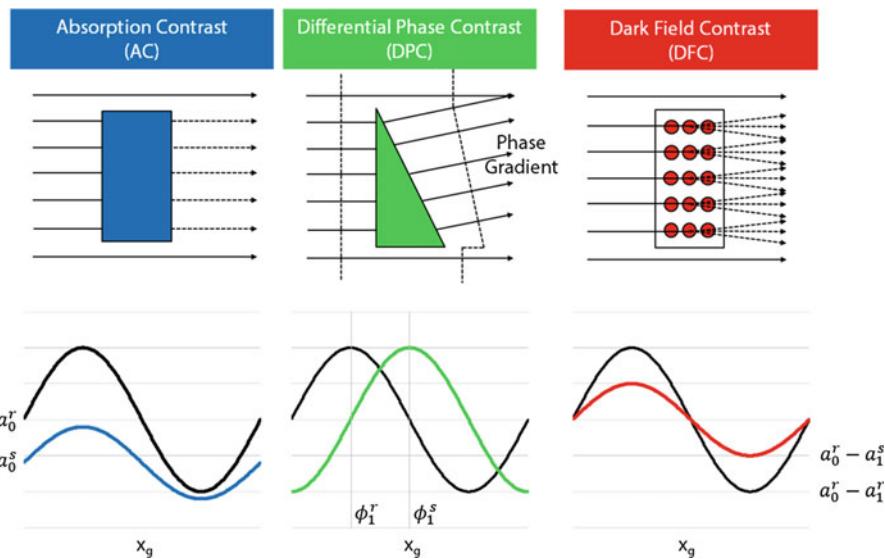
Figure 39 shows the principle of a Talbot-Lau grating interferometer XCT: A source grating ( $G_0$ ) acts as an array of line sources, ensuring a sufficient transverse coherence length, while a diffractive grating ( $G_1$ ) causes phase modulation of the incoming X-rays resulting in an interference pattern. As the fringe pattern cannot be resolved directly with conventional X-ray detectors, an absorption grating ( $G_2$ ) is placed right in front of the detector.  $G_1$  and  $G_2$  are setting up the interferometer. The source grating  $G_0$  is the phase-stepping grating; it is used to scan transversely across the intensity pattern. The result is a sinusoidal intensity modulation of the signal.



**Fig. 39** Principle of a Talbot-Lau grating interferometer XCT setup with source grating G0, phase grating G1, and absorption grating G2 (left). For measuring AC, DPC, and DFC, the phase source grating G0 measures the intensity at different grating positions. For XCT the object is turned

Rotating the sample stepwise in between the acquisition of the phase stepping curves allows the tomographic operation of such a setup. Placing a sample within optical beam axis will attenuate, refract, and scatter X-rays and thus perturbing the periodic intensity modulations. A Fourier analysis of the intensity modulations of each detector pixel is used to simultaneously extract co-registered AC, DPC, and DFC. The image formation of the different contrast modalities in a Talbot-Lau grating interferometer XCT is shown in Fig. 40. The attenuation of the X-ray beam results in a reduction of the incident intensity. A purely phase-shifting sample (in this case using an idealized wedge) leads to a lateral shift of the interference pattern, and results in an interferometer phase-shift of the measured phase-stepping curve. An idealized, scattering-only sample leads to a reduction of the oscillation height and to a reduction of the visibility measured in the dark-field image. Talbot-Lau grating interferometry is an attractive method for practical applications since commercial desktop XCT systems are already available.

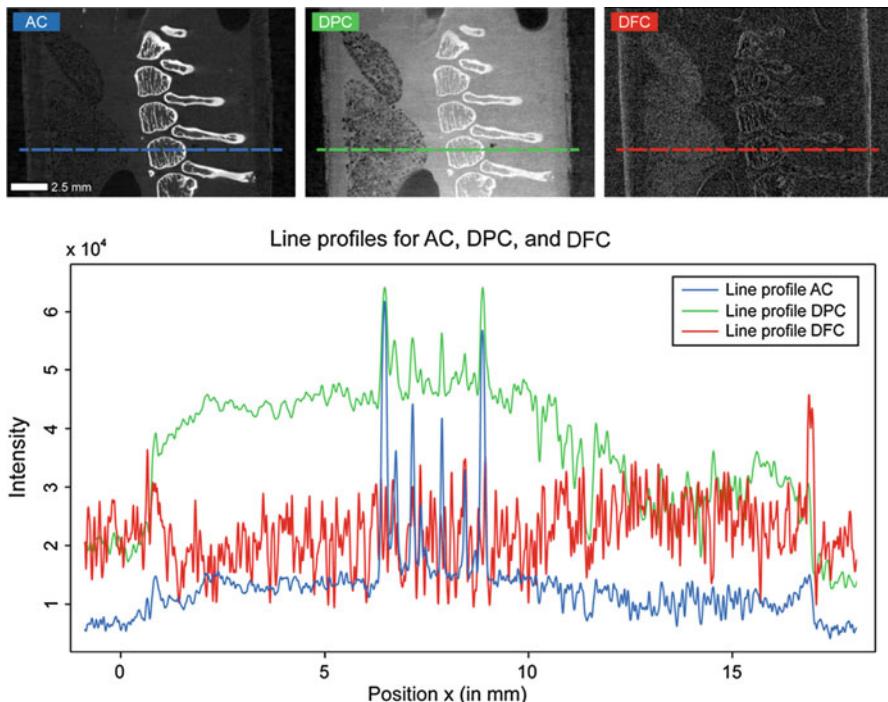
In Fig. 41, cross-sectional XCT images and corresponding line profiles of a mole are shown. In the left image the absorption contrast is depicted, phase contrast in the center and dark field scatter contrast in the right image. The breast of the mole is visible including the lung and the backbone. Within the absorption data the bone features the highest contrast and the lung is almost invisible. In the dark field contrast image, the lung and the bone show a similar contrast since the dark field signal is caused by microscopic inhomogeneities in the mole represented by pores and fibers. In the bottom image of Fig. 41, the gray value intensity profiles of the AC, DPC, and DFC images along the dashed lines of the samples (see cross sections) are shown.



**Fig. 40** TLGI provides three complementary modalities: Attenuation contrast (AC) due to absorption of the prevalent material, differential phase contrast (DPC) due to refraction and dark-field contrast (DFC) due to scattering. Attenuation of the X-ray beam results in a reduction of the intensity. A purely phase-shifting sample leads to a lateral shift of the interference pattern. An idealized, scattering-only sample leads to a reduction of the oscillation height. (Adapted from Malecki 2013)

In Fig. 42 AC, DFC, and DPC axial slice images of a carbon fiber-reinforced polymeric (CFRP) sample are shown, which have been acquired with different XCT systems. In the AC images, the fiber bundles cannot be distinguished from the epoxy matrix. They show limited contrast in the DPC images. However, the fiber bundles are heterogeneous by nature and thus lead to a strong scattering signal. Consequently, the bundles parallel to the cutting plane appear clearly in the respective cross-sectional image of the scattering coefficient in the form of bright stripes. Near the surface there is a copper mesh placed in the specimen for lightning strike protection, which is causing severe artifacts in the AC data. The DPC image shows almost no streaking artifacts and features an increased contrast between pores and matrix. This behavior can be explained by the fact that DPC shows a reduced dynamic range and gray value range as compared to AC. Pores located close or in the layer of the copper mesh are now visible in the corresponding DPC axial slice image. These pores have been hidden in AC by strong artifacts. The DFC image reveals individual carbon fiber bundles along a certain orientation.

In the next example the 3D structure of carbon fiber bundles (yarn) is analyzed. Figure 43 shows 3D representations of a CFRP sample that was scanned by XCT absorption contrast (left image) and TLGI-scattering XCT contrast (right image). The “plain-weave” CFRP structure is hardly recognizable in the AC data but can be clearly seen in the fused 0° and 90° DFC datasets, and the 3D yarn structure may be



**Fig. 41** Cross-sectional XCT pictures of a mole with absorption contrast (left), phase contrast (center), and dark field scatter contrast (right). The diagram in the bottom shows line profiles through the respective sections in absorption (blue), phase (green), and dark-field contrast (red). The XCT data were acquired with a Talbot-Lau desktop interferometer. Voxel size was  $11.4 \mu\text{m}^3$ . (Adapted from Kastner et al. 2017a)

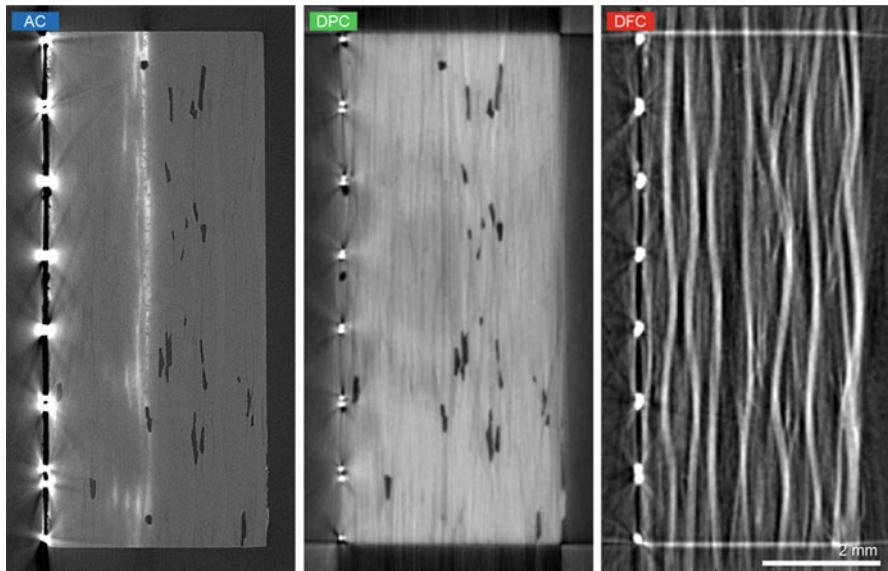
analyzed quantitatively. This example demonstrates the orientation dependence of DFC and the usefulness of DFC for fiber bundle analysis.

As reported above TLGI-XCT shows high potential for practical applications although it is not yet very common. In addition, there are several promising developments on the basis of TLGI such as vector radiography, tensor tomography, etc. (Prade et al. 2017).

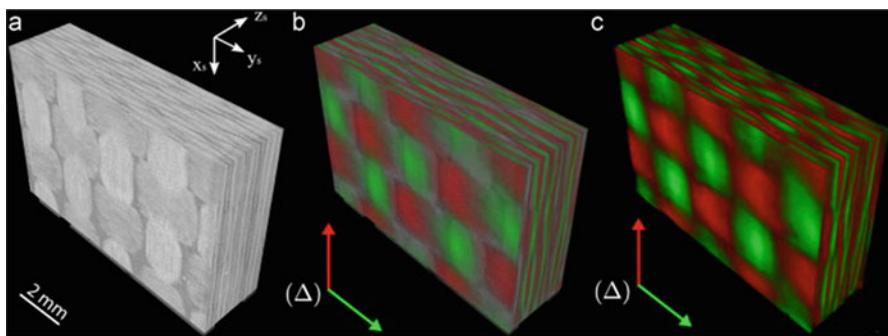
## Trends and Summary

There is a large number of further trends and new developments which were not discussed above. These trends comprise aspects such as:

- New X-ray tubes, in particular nano-focus tubes, high-energy tubes up to 800 keV tube, liquid-metal-jet X-ray sources with high power, etc. New detectors, in particular photon counting detectors



**Fig. 42** Cross-sectional images of attenuation contrast (AC data), differential phase contrast (DPC data) and dark-field image (DFC data) of a carbon fiber-reinforced polymeric (CFRP) sample with a copper mesh near the surface. The voxel size was  $22.8 \mu\text{m}^3$



**Fig. 43** Three-dimensional XCT data of a CFRP laminate sample generated from AC **(a)**, overlay between AC image and merged scatter dark-field datasets **(b)** and merged scatter dark-field datasets **(c)** images. The direction of the rotation axis ( $\Delta$ ), and accordingly the grating lines, is indicated for both sample orientations in red and green, respectively.  $0^\circ$  (green) and  $90^\circ$  (red) scans (Revol et al. 2013)

- New and improved reconstruction techniques as well as scanning strategies (e.g., algebraic reconstruction, statistical reconstruction methods, discrete tomography, etc.)
- Improved and new methods for measurement artifact reduction
- Improved and new methods for more accurate XCT simulation

- Elemental XCT/spectral imaging by application of energy-dispersive single photon counting detectors, K-edge imaging, and X-ray fluorescence tomography
- Multimodal imaging by combination of X-ray XCT with complementary imaging methods
- Laboratory diffraction contrast tomography enabling crystallographic imaging in 3D dimensions with lab XCT system (McDonald et al. 2017)
- Higher resolution, better accuracy – more quantitative values, faster scanning (e.g., for in- or at-line XCT), scanning of larger parts, etc.

Summarizing, X-ray computed tomography is still an emerging technology with many applications which are yet unknown. Some of the upcoming applications may be addressed by XCT technology, which is already available, while others will require advancements of XCT. In this review we have reported on different aspects and trends of XCT for nonmedical applications focusing on:

- Principles of XCT, measurement artifacts and their correction and a summary of different XCT methods
- Applications of X-ray CT for materials characterization and metrology
- Advanced XCT Data processing, analysis and visualization methods, Quantitative XCT: XCT as a measurement tool for extracting characteristics of materials and components as well as their distributions, XCT as the basis for materials simulation
- 4DXCT or in situ XCT: XCT as a tool to study processes, damage growth, etc.
- Phase-contrast XCT methods such as Propagation-based phase-contrast XCT and Talbot-Lau grating interferometer XCT.

**Acknowledgments** This work was supported by the project “Multimodal and in-situ characterization of inhomogeneous materials” (MiCi) of the federal government of Upper Austria and the European Regional Development Fund (EFRE) in the framework of the EU program IWB2020. The research leading to these results has also received funding from the FFG Bridge Early Stage, project no. 851249 (“Advanced multimodal data analysis and visualization of composites based on grating interferometer micro-CT data (ADAM)”) as well as from the FWF-FWO 2016 Lead Agency Call for Joint Projects, project no. I3261-N36/S004217 N (“Quantitative X-ray tomography of advanced polymer composites”).

---

## References

- Adams R, Bischof L (1994) Seeded region growing. *IEEE Trans Pattern Anal Mach Intell* 16:641–647. <https://doi.org/10.1109/34.295913.3>
- Amirkhanov A (2012) Visualization of Industrial 3DXCT Data, PhD Thesis, TU Vienna
- Amirkhanov A, Heinzl C, Reiter M, Gröller E (2010) Visual Optimality and Stability Analysis of 3DCT Scan Positions. *IEEE Trans Vis Comput Graph* 16:1477–1487
- Amirkhanov A, Heinzl C, Kastner J, Gröller E, Fuzzy CT (2013) Metrology: dimensional measurements on uncertain data. In: SCCG proceedings (digital library) Smolenice castle, Slovakia, p 8

- Amirkhanov A, Amirkhanov A, Salaberger D, Kastner J, Gröller M, Heinzl C (2016) Visual analysis of defects in glass fiber reinforced polymers for 4DCT interrupted in-situ tests. *Comput Graphics Forum* 35:201–210
- Banhart J (ed) (2008) Advanced tomographic methods in materials research and engineering. Research Oxford University Press, Oxford
- Bartscher M, Neuschaefer-Rube U, Illemann J, Borges de Oliveira F, Stolfi A, Carmignato S (2018) Qualification and testing of CT systems. In: Carmignato S, De Wulf W, Bartscher M (eds) Industrial X-ray computed tomography. Springer, Cham
- Baruchel J, Buffiere JY, Maire E, Peix G (eds) (2000) X-ray tomography in material science. Hermes Science Publications, Paris
- Beucher S, Lantuejoul C (1979) Use of watersheds in contour detection. In: International workshop on image processing: real-time edge and motion detection/estimation, Rennes, Sept 1979
- BIPM JCGM 100 (2008) Evaluation of measurement data – guide to the expression of uncertainty in measurement. International Organisation for Standardisation, Geneva
- Buffiere J-Y, Maire E, Adrien J, Masse J-P, Boller E (2010) In-situ experiments with X ray tomography: an attractive tool for experimental mechanics. *Exp Mech* 50:289–305
- Buzug TM (2008) Computed tomography: from photon statistics to modern Cone-Beam CT. Springer, Berlin/Heidelberg
- Carmignato S, Dewulf W, Leach R (eds) (2018) Industrial X-ray computed tomography. Springer, Berlin/Heidelberg
- De Chiffre L, Carmignato S, Kruth JP, Schmitt R, Weckenmann A (2014) Industrial applications of computed tomography. *CIRP Ann Manuf Technol* 63:655–677
- Feldkamp LA, Davis LC, Kress JW (1984) Practical cone beam algorithm. *J Opt Soc Am* 6:612–619
- Gusenbauer C, Reiter M, Salaberger D, Kastner J (2016) Comparison of metal artefact reduction algorithms from medicine applied to industrial XCT applications. In: Proceedings 19th World conference on non-destructive testing 2016 (WCNDT 2016). [www.ndt.net](http://www.ndt.net)
- Hanke R, Fuchs T, Uhlmann N (2008) X-ray based methods for non-destructive testing and material characterization. *Nucl Inst Methods Phys Res A* 59:14–18
- Harrer B, Kastner J (2011) Cone Beam CT for non-destructive testing Fabrication and Characterization in the Micro-Nano Range: new Trends for two and three dimensional Structures, ‘X-ray Microtomography: characterization of structures and defect analysis’. Springer, Heidelberg, pp 119–150
- Heinzel C, Stappen S (2017) STAR: visual Computing in Materials Science. *Comput Graphics Forum* 36(3):647–666
- Heinzel C, Kastner J, Gröller E (2007) Surface extraction from multi-material components for metrology using dual energy CT. *IEEE Trans Vis Comput Graph* 13:1520–1528
- Heinzel C, Amirkhanov A, Kastner J (2018) Processing, analysis and visualization of CT data. In: Carmignato S, De Wulf W, Bartscher M (eds) Industrial X-ray computed tomography. Springer, Cham
- Herman GT (1979) Correction for Beam Hardening in computed tomography. *Phys Med Biol* 24:81–106
- Hermanek P, Carmignato S (2017) Porosity measurements by X-ray computed tomography: accuracy evaluation using a calibrated object. *Precis Eng* 49:377–387
- Hsieh J (2003) Computed tomography, principles, design, artifacts and recent advances. In: SPIE The International Society for Optical Engineering, Bellingham
- Janssens E, Pereira LF, De Beenhouwer J, Tsang R, Van Dael M, Verboven P, Nicolai B, Sijbers J (2016) Fast inline inspection by Neural Network Based Filtered Backprojection: application to apple inspection. *Case Stud Nondestruct Test Eval* 6:14–20. <https://doi.org/10.1016/j.csndt.2016.03.003>
- Jensen T, Bech M, Bunk O, Donath T, David C, Feidenhans'l R, Pfeiffer F (2010) Directional x-ray dark-field imaging. *Phys Med Biol* 55:3317–3323
- Joseph PM, Spital RD (1982) The effects of scatter in x-ray computed tomography. *Med Phys* 9:464–472

- Kasperl S (2005) Qualitätsverbesserungen durch referenzfreie Artefaktreduzierung und Oberflächennormierung in der industriellen 3D-Computertomographie. PhD Thesis, Technische Fakultät der Universität Erlangen, Nürnberg
- Kastner J (2016) Proceedings of 6th conference on industrial computed tomography (iCT2016) 2016, 9–12 Feb 2016, Wels. [www.ndt.net](http://www.ndt.net)
- Kastner J, Harrer B, Requena G, Brunke O (2010a) A comparative study of high resolution cone beam X-ray tomography and synchrotron tomography applied to Fe- and Al-alloys. *NDT&E Int* 43:599–605
- Kastner J, Plank B, Salaberger D, Sekelja J (2010b) Defect and porosity determination of fiber reinforced polymers by X-ray computed tomography. In: 2nd international symposium on NDT in Aerospace. [www.ndt.net](http://www.ndt.net)
- Kastner J, Salaberger D, Plank B (2012a) High resolution X-ray computed tomography of fibre and particle filled polymers. In: Proceedings of 12th World conference on non-destructive testing, Durban. [www.ndt.net](http://www.ndt.net)
- Kastner J, Salaberger D and Plank B (2012b) Microstructure characterization of high-strength Al-alloys by high resolution X-ray computed tomography. In: Proceedings of 12th World conference on non-destructive testing
- Kastner J, Harrer B, Requena G (2012c) Non-destructive characterization of polymers and Al-alloys by poly-chromatic cone-beam phase contrast tomography. *Mater Charact* 64:79–87
- Kastner J, Zaunschirm S, Baumgartner S, Requena G, Pinto H, Garcés G (2014) 2D-microstructure characterization of thermo-mechanically treated Mg-alloys by high resolution X-ray computed tomography. In: 11th European conference on non-destructive testing (ECNDT 2014), October 6–10, Prague
- Kastner J, Heinzl C, Plank B, Salaberger D, Gusenbauer C, Senck S (2017a) New X-ray computed tomography methods for research and industry. In: 7th conference on industrial computed Tomography (iCT2017), Leuven
- Kastner J, Salaberger D, Heinzl C, Gusenbauer C, Rao G (2017b) High resolution X-ray computed tomography for non-destructive characterization and in-situ investigations. In: 15th Asia Pacific conference for non-destructive testing (APCNDT2017), Singapore
- Krumm M, Kasperl K, Franz M (2008) Reducing Non-linear artifacts of multimaterial objects in industrial 3D computed tomography. *NDT&E Int* 41:242–251
- Kruth JP, Bartscher M, Carmignato S, Schmitt R, De Chiffre L, Weckenmann A (2011) Computed tomography for dimensional metrology. *CIRP Ann Manuf Technol* 60:821–842
- Lambert JH (1760) *Photometria, sive de mensura et gradibus luminis colorum et umbrae. Sumptibus Viduae Eberhardi Klett*, Leipzig
- Landis EN, Keane DT (2010) X-ray microtomography. *Mater Charact* 61:1305–1316
- Liu Y, Straumit I, Vasiukov D, Lomov SV, Panier S (2017) Prediction of linear and non-linear behavior of 3D woven composite using mesoscopic voxel models recon-structed from X-ray micro-tomography. *Compos Struct* 179:568–579
- Maire E, Withers PJ (2014) Quantitative X-ray tomography. *Int Mater Rev* 59:1–43
- Malecki A D (2013) X-ray tensor tomography from two-dimensional directional X-ray dark-field imaging to three dimensions, PhD Thesis, TU Munich
- McDonald AA, Holzner C, Lauridsen EM, Reischig P, Merkle AP, Withers PJ (2017) Microstructural evolution during sintering of copper particles studied by laboratory diffraction contrast tomography (LabDCT). *Sci Rep* 7:5251. <https://doi.org/10.1038/s41598-017-04742-1>
- Meyer E, Raupach R, Lell M, Schmidt B, Kachelrieß M (2010) Normalized metal artifact reduction (NMAR) in computed tomography. *Med Phys* 37:5482–5493
- Millner MR, McDavid WD, Waggener RG, Dennis MJ, Payne WH, Sank VJ (1979) Extraction of information from CT scans at different energies. *Med Phys* 6:70–71. <https://doi.org/10.1111/1.594555>
- Müller P (2012) Coordinate metrology by traceable computed tomography, pavel müller, PhD thesis, Technical University of Denmark
- Oberpeilsteiner S, Salaberger D, Reiter T (2014) Coupling of in-situ-CT with virtual testing by FEM of short fiber reinforced materials. In: Proceeding ICT2014, 25–28 Feb 2014, Wels

- Oh W, Lindquist WB (1999) Image thresholding by indicator kriging. *IEEE Trans Pattern Anal Mach Intell* 21:590–602
- Oster R (1999) Computed tomography as a non-destructive test method for Fiber Main Rotor Blades in development, series and maintenance. In: Proceedings of international symposium on digital industrial radiology and computed tomography Berlin (DGZfP-Berichtsband 67-CD), [www.dgzfp.de](http://www.dgzfp.de). (DGZfP-Berichtsband 67-CD)
- Paganin D, Mayo SC, Gureyev TE, Miller PR, Wilkins SW (2002) Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object. *J Microsc* 206:33–40. <https://doi.org/10.1046/j.1365-2818.2002.01010.x>
- Pereira L, Janssens E, Cavalcanti G, Tsang IR, Van Dael M, Verboven P, Nicolai B, Sijbers J (2017) Inline discrete tomography system. *Comput Electron Agric* 138:117–126. <https://doi.org/10.1016/j.compag.2017.04.010>
- Pfeiffer F, Weitkamp T, Bunk O, David C (2006) Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources. *Nat Phys* 2:258–261
- Plank B, Kastner J, Schneider R, Busch R (2009) Charakterisierung von Porositäten in Sintermetallen mit 3D-CT Anwendungsbeispiel selektives Lasersintern. In: Proceedings of DGZfP annual conference. DGZfP, Münster, pp 810–816
- Prade F, Schaff F, Senck S, Meyer P, Mohr J, Kastner J, Pfeiffer F (2017) Nondestructive characterization of fiber orientation in short fiber reinforced polymer composites with X-ray vector radiography. *NDT&E Int* 86:65–72
- Reh A, Plank B, Kastner J, Gröller E, Heinzl C (2012) Porosity maps: interactive exploration and visual analysis of porosity in carbon fiber reinforced polymers using X-ray computed tomography. *Comput Graph Forum* 31(3):1185–1194
- Reims N, Schoen T, Boehnel M, Sukowski F, Firsching M (2014) Strategies for efficient scanning and reconstruction methods on very large objects with high-energy x-ray computed tomography. *SPIE Optical Engineering & Applications*:921209–921209
- Reiter M, Erler M, Kuhn C, Gusenbauer C, Kastner J (2016) SimCT: a simulation tool for X-ray imaging. In: Proceedings 6th conference on industrial computed tomography (iCT2016) 2016, 9–12 February 2016, Wels. [www.ndt.net](http://www.ndt.net)
- Requena G, Cloetens P, Altendorfer W, Poletti C, Tolnai D, Warchomicka F, Degischer HP (2009) Sub-micrometer synchrotron tomography of multiphase metals using Kirkpatrick–Baez optics. *Scr Mater* 61:760–763. <https://doi.org/10.1016/j.scriptamat.2009.06.025> ISSN 1359-6462
- Revol V, Plank B, Kaufmann R, Kastner J, Kottler C, Neels A (2013) Laminate fiber structure characterization of carbon fiber-reinforced polymers by X-ray scatter dark field imaging with a grating interferometer. *NDT&E Int* 58:64–71
- Russ JC (2002) The image processing handbook. CRC Press LLC, Boca Raton
- Salaberger D, Kannappan KA, Kastner J, Reusssner J, Auinger T (2011) CT Data evaluation of fiber reinforced polymers to determine fiber length distribution. *Int Polym Process* 3:283–291
- Salvo L, Suérya M, Marmottant A, Limodin N, Bernard D (2010) 3D imaging in material science: application of X-ray tomography. *C R Phys* 11:641–649
- Schladitz K (2011) Quantitative micro-CT. *J Microsc* 243:111–117
- Schmitt R, Niggemann C (2010) Uncertainty in measurement for X-Ray computer tomography using calibrated work pieces. *Meas Sci Technol* 21:1–9
- Sezgin M, Sankur B (2004) Survey over image thresholding techniques and quantitative performance evaluation. *J Electron Imaging* 13(1):146–168
- Sheppard AP, Sok RM, Averdunk H (2004) Techniques for image enhancement and segmentation of tomographic images of porous materials. *Phys A* 339:145–151
- Straumit S, Hahn C, Winterstein E, Plank B, Lomov SV, Wevers M (2016) Computation of permeability of a non-crimp carbon textile reinforcement based on X-ray computed tomography images open overlay panel. *Compos A: Appl Sci Manuf* 81:289–295
- Tuy-Smith K (1983) An inversion formula for cone-beam reconstruction. *SIAM J Appl Math* 43:546–552
- Ueda R, Kudo H, Dong J (2017) Applications of compressed sensing image reconstruction to sparse view phase tomography. In: Proceedings of SPIE 10391, developments in X-ray tomography XI, 103910H, 3 Oct 2017. <https://doi.org/10.1117/12.2273691>

- Wiegert J (2007) Scattered radiation in cone-beam computed tomography: analysis, quantification and compensation. PhD Thesis, RWTH Aachen, Aachen
- Williams J, Yazzie KE, Padilla E, Chawla N, Xiao X, De Carlo F (2013) Understanding fatigue crack growth in aluminum alloys by in-situ X-ray synchrotron tomography. *Int J Fatigue* 57:79–85
- Yashiro W, Terui Y, Kawabata K, Momose A (2010) On the origin of visibility contrast in x-ray Talbot interferometry. *Opt Express* 18:16890–16901
- Zhigeng P, Jianfeng L (2007) A bayes-based region-growing algorithm for medical image segmentation. *Comput Sci Eng* 9:32–38
- Zhou S-A, Brahme A (2008) Development of phase-contrast X-ray imaging techniques and potential medical applications. *Phys Med* 24:129–148
- Zhou J, Maisl M, Reiter H, Arnold W (1996) Computed laminography for materials testing. *Appl Phys Lett* 68:3500–3502



# 3D X-Ray Tomography: Basics and Latest Developments

32

Theobald O. J. Fuchs and Randolph Hanke

## Contents

Introduction .....	1168
Physical Basis of Industrial X-Ray Technology .....	1169
Attenuation of X-Ray Radiation .....	1169
Generation of X-Rays .....	1170
Detection of X-Ray Radiation: Image Acquisition Technologies .....	1172
Projection Method .....	1173
Process for Sectional Imaging with X-Ray Radiation .....	1174
Micro-Und Nano-CT .....	1175
New CT-Methods Based on Iterative Reconstruction Methods .....	1175
Outlook on the Future Development of Industrial X-Ray Imaging .....	1178
References .....	1179

## Abstract

In the following, the basic principles of X-ray physics are discussed which includes generation and detection of X-rays and the acquisition of X-ray projection images. Further on, the process of computer-assisted sectional image calculation is briefly introduced and the latest developments in the field are mentioned. Additionally, particular issues of micro- and nano-scale X-ray Computed Tomography are described. Finally, we attempt to look forward into the upcoming future of industrial X-ray imaging systems which most probably will evolve to cognitive sensor networks by applying advanced machine-learning technologies.

---

T. O. J. Fuchs

Fraunhofer EZRT, Fürth, Germany

e-mail: [theobald.fuchs@iis.fraunhofer.de](mailto:theobald.fuchs@iis.fraunhofer.de)

R. Hanke (✉)

Fraunhofer IZFP, Saarbrücken, Germany

e-mail: [randolf.hanke@iis.fraunhofer.de](mailto:randolf.hanke@iis.fraunhofer.de)

## Introduction

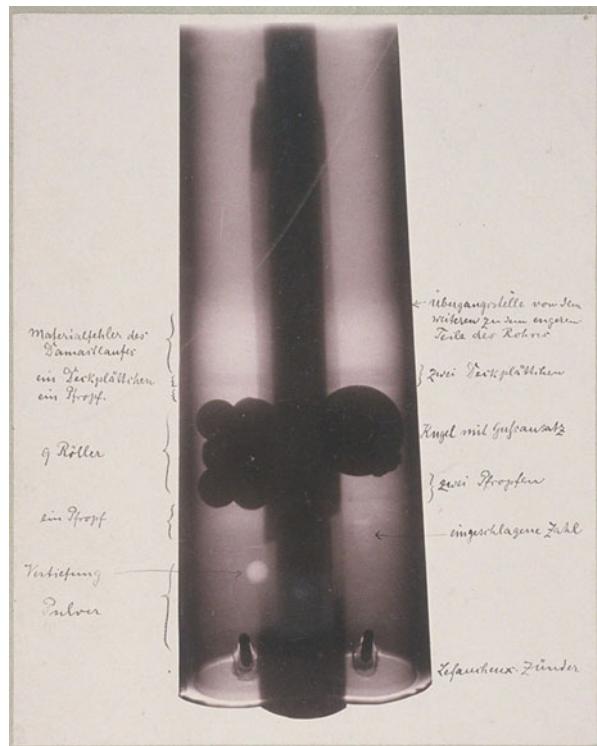
At the beginning of the twenty-first Century, X-ray inspection is one of the most efficient tools used for nondestructive testing and evaluation in industrial production and research.

The range of established techniques encompasses digital radioscopy and computer tomography as well as small angle- and back-scattering. Newly developed materials, such as those used in aircraft construction, require an increasing use of X-ray technology, and at the same time X-ray imaging is still making significant technical and scientific progress in the first two decades of the new millennium.

In the following, we will first take a closer look at the basic principles of industrial X-ray imaging. Then examples of current applications will be given and the resulting challenges expounded. Finally, the corresponding potential for further development will be presented.

The basic demands and challenges, which still play a major role in the further development of nondestructive testing using X-rays, are still of the same nature as those experienced by Wilhelm C. Röntgen in 1895 during his attempt to use the recently discovered X-rays to examine his hunting rifle (Fig. 1):

**Fig. 1** X-ray image of Röntgen's Hunting rifle with his original annotation hand written on the film. (Source: German X-Ray-Museum in Remscheid, Wilhelm C. Röntgen's birthplace)



- The exponential attenuation of the rays in materials causes considerable differences in the intensities measured behind a specimen.
- Due to the projective nature of the imaging geometry, object structures overlap in the image in the direction of the radiography (“projection technique”).
- In practice, to achieve an acceptable proportion of contrast to background noise in the image, in particular for very large and/or dense objects, long exposure times are often necessary.
- In many cases, the contrast is noticeably reduced due to the diffuse background from scattered radiation.
- The significant attenuation of X-ray radiation through metals compared to human tissue and organs (with the exception of the skeleton).

Since the late 1990s, for the industrial examination using X-rays, two techniques are used on a large scale: On the one hand, the projection method, in which two-dimensional radiographic images are generated, and, on the other hand, computed tomography with which three-dimensional volume data sets are obtained, representing the spatial distribution of the material density within the object being examined.

Along with these two most important imaging processes for X-ray based defect-monitoring in industrial development and production (Fuchs and Hanke 2008), there are other processes which are only mentioned here for the sake of completeness: X-ray fluorescence analysis and diffraction, which serve to analyze the fine structure of the object being examined as well as the emerging technology of phase-contrast-based imaging, which is, however, limited to relatively small and light objects due to the low X-ray energy and beam coherence used.

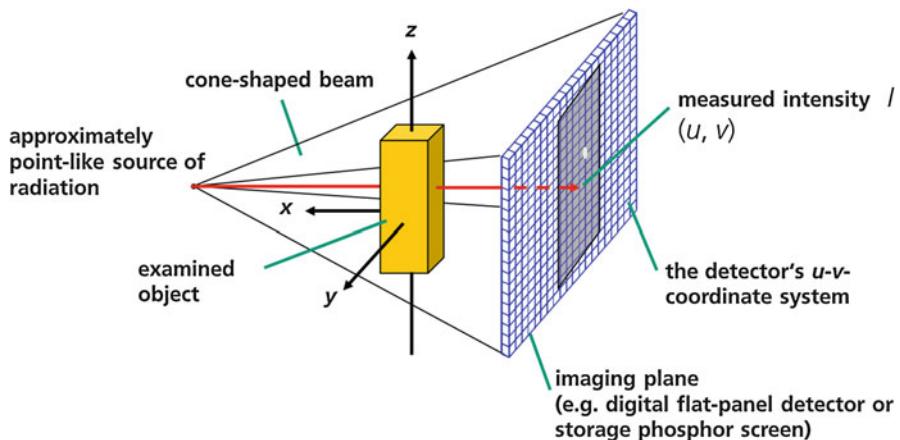
---

## Physical Basis of Industrial X-Ray Technology

### Attenuation of X-Ray Radiation

In radiology as well as in computed tomography which is derived from the projective X-ray imaging, the special properties of electromagnetic radiation in an energy range of between 5 keV and 10 MeV are used in order to pass more or less unattenuated through solid (or liquid) substances. The attenuation depends on the following three factors: Composition of the material from its chemical elements, radiation energy, and the thickness of the material to be penetrated. The spatial distribution of the radiation which occurs after passing through and exiting from an object provides information about the structure and composition of the interior of the object (Fig. 2).

The attenuation of the X-ray radiation in a material increases quadratically with the atomic number of the material. Heavier materials, such as steel ( $Z = 26$ ), absorb a disproportionately larger amount of radiation than materials which have a lower atomic number (such as aluminum which has the nucleus charge value of 13). Thicker walls also increase the absorption.



**Fig. 2** Speaking precisely with respect to physics, a radiographic image is the result of a measurement of spatial distribution of the output intensity of X-ray radiation, after it has passed through an object (typically in either a solid or a liquid physical state and described in a well-defined, e.g.,  $x$ - $y$ - $z$  Cartesian, coordinate system)

The X-ray radiation which remains after passing through the object follows the so-called *Lambert-Beer's law* of attenuation. This law allows us to predict the expected average transmitted intensity. According to this, the incoming intensity  $I_0$  decreases exponentially along the attenuation path  $d$  through the object:

$$I = I_0 e^{-\mu d}$$

The linear attenuation coefficient (also known as absorption coefficient)  $\mu$  is specific to the material being examined and results from the product of the density  $\rho$  and the sum of cross-sections which again depend on energy and material (essentially the photo-effect and the Compton scattering).  $\mu$  increases with increasingly higher atomic number and thus describes the abovementioned behavior of the absorption (Hubbell 1982; Hubbell and Seltzer 1989).

## Generation of X-Rays

In an industrial setting, electrically operated X-ray tubes are most often used, which mainly generate *Bremsstrahlung*. Accelerated electrons, which are focused on a so-called target made of metal (e.g., a disk, the size of a small coin and made of tungsten) release their kinetic energy upon impact, whereby deceleration and ionization processes play a role. During these processes, short wave X-ray light (amongst others) is radiated. The required electrons are emitted either from a hot cathode, i.e., a helical coil (also known as filament) or an electrically heated thin metal layer and accelerated by means of an electric field in the direction of the target. The electric field is generated by a high-voltage generator. The typically used voltage range lies

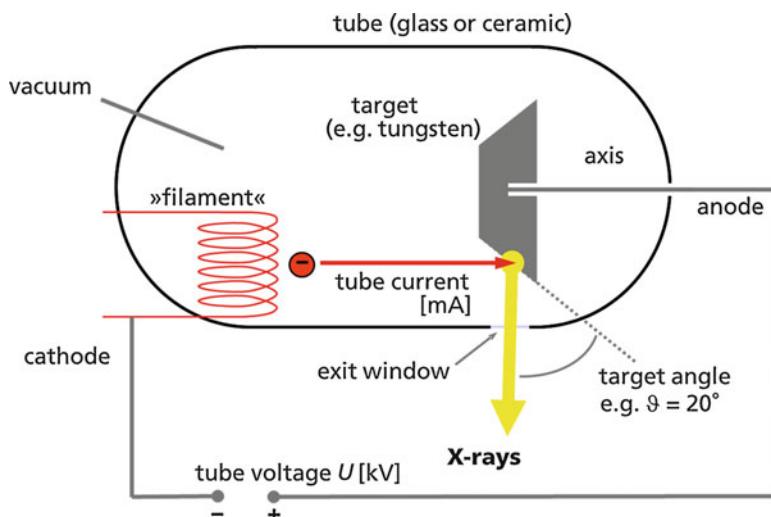
between a few kilovolts up to several megavolts, whereby the value of the high voltage applied determines the maximum photon energy of the resulting X-ray radiation.

In the target, which is generally made of a technical metal with as high a melting point as possible, the incident electrons lose their energy through scattering in a Coulomb potential of the atom nuclei or through ionization of the atoms. In both cases, the X-ray quanta generated can be distinguished correspondingly as Bremsstrahlung or characteristic radiation. However, more than 99% of the electron energy deposited in the metal is converted to heat and less than 1% to X-ray photons, which are then available as a useful beam for X-ray imaging.

The parameters which determine the radiation's intensity and energy at the output of an X-ray tube correspond to the emission flow in the cathode and the acceleration voltage between cathode and target, which in most cases is formed by the anode of a vacuum tube (Fig. 3).

As previously mentioned, the process of X-ray beam generation is very inefficient. In order to prevent the target from melting due to high heat input, occasionally so-called rotating-anode tubes were designed on which the thermal energy is distributed over a larger surface area. However, in any case, the electrical tube power must be limited. The danger of thermal damage to the X-ray tubes causes limitations in the measuring time or vice versa the object size.

Another quite new possibility to increase the photon flux of X-ray tubes is the Liquid Metal Jet technology. The problem of heat generation and target melting is solved by replacing solid-metal anodes by liquid metal jets made out of gallium alloys (Hemberg et al. 2003; Otendal 2006).



**Fig. 3** Schematic longitudinal section through an X-ray tube equipped with a rotating target on which the heat, introduced when the high energetic electron beam is rapidly decelerated, can be better distributed

This constraint has a particular influence on radiosscopic imaging which is based on the method of direct magnification: For a good spatial resolution in the object to be examined, which is a decisive factor for the recognition of smaller details, a smaller X-ray focal spot is necessary. However, a size reduction of the focal spot also necessitates a greater reduction in the tube power in order to prevent the previously mentioned melting of the target. The subsequent reduced primary intensity in turn causes an increase in exposure times and total measuring times, which is required in order to obtain a sufficiently high signal-to-noise-ratio for reliable fault detection.

In addition to this, the distribution of the photon energies is critical for the penetration ability of the X-ray beam. This also influences the linear attenuation coefficient  $\mu$  of a material. If denser and heavier materials are to be irradiated, accordingly higher acceleration voltages are required.

---

## Detection of X-Ray Radiation: Image Acquisition Technologies

There are two principles which are most frequently used for detection of X-ray radiation in radiosscopic imaging (cf. also Sukowski et al. 2009; Scholz et al. 2009; Salamon et al. 2008b, 2009).

Mostly, the semi-conductor materials silicon (Si), cadmium-telluride (CdTe), cadmium-zinc-telluride (CZT), or gallium-arsenide (GaAs) are used as sensor materials for *directly converting detectors*. The incident X-ray photons generate free charge carriers directly during absorption in a layer of the semi-conductor. These are collected on pixelated electrodes. The charge measured there correlates with the energy deposited by a photon. The principle of direct conversion is the basis for two important features of this kind of detectors:

- **Energy Resolution:** Detectors with readout electronics with more than one channel can differentiate between the detected X-ray photons with regard to their energy differences.
- **Counting of Individual Photons:** Due to the possibility of suppressing the electronic noise, a directly converting detector can be exposed over a long period of time and in this way single X-ray photons can be detected.

The use of *indirect detection of X-ray radiation* is more common and standard in an industrial setting (also known as “indirectly converting” or “integrating” sensors). In this case, the incoming radiation first passes through a special foil (the scintillator), in which the X-ray photons cause a visible scintillation flash, whose brightness is proportional to the intensity and energy of the absorbed radiation. Behind this scintillating layer, a camera or a photodiode array collects the light and converts it into grey levels of a digital X-ray image. The indirect detection principle is used in most modern digital detector arrays (“flat panel detectors”).

Whereas the strengths of directly converting detectors lie especially in the area of low energy applications (photon energies from around 10 keV to 50 keV) for the detection of very low intensities and spectral imaging, indirectly converting X-ray

detectors are recommended for the region above 60 keV as well as for sensors providing large sensitive areas.

A further important criterion, which is crucial when selecting and characterizing a detector, is its dynamic range, i.e., the ratio of digitization depth to background noise. The dynamic range actually determines the obtainable contrast resolution which describes the ability of a detector system to depict differences in wall thickness. Thanks to a nominal digitization depth of up to 16 bit with around 2 or 3 bits of dark noise level, modern systems such as flat panel detectors based on amorphous silicon (a-Si) can detect thickness variations of less than 1% (e.g., detection of a 200 µm fault when wall thickness is 20 mm).

## Projection Method

In radioscopy, a projection image of an object is created using X-ray radiation according to the same principle as in geometric optics: The radiation originating from a point source penetrates the object and in doing so is attenuated according to the shape and inner structure of the object. The silhouette is magnified geometrically directly, i.e., during the process of measuring according to the relative distances from source, object, and image plane. The distribution of attenuated radiation is observed on the side opposite the source of radiation (cf. Fig. 2).

In the case of X-ray radiation, the projection image provides a measure at every point for the integrated attenuation capability of the material along a beam through the object. In the past, the recording of the attenuated radiation intensity was carried out using film. Nowadays it is carried out increasingly often using electronic digital detector arrays. The resolution in the resulting projection image is determined by the so-called geometrical blurring and by possible limitations of the imaging system (film or flat panel detector).

When the magnification factor is  $V$  and the expansion of the radiation source in a particular direction,  $f$ , are known, the geometrical blurring in the specified direction can be calculated by

$$u_{\text{geo}} = f \times (V - 1)$$

The spatial sampling defined by the pixel-to-pixel distance  $p$  of a detector also contributes to the blurring of a projection image. It can effectively be calculated in the object with the reciprocal magnification factor, i.e.,  $p / V$ .

For images, which are carried out with lower or no magnification at all (i.e.,  $V \approx 1$ ), the resulting resolution is thus determined by the detector. When higher magnification factors are used, the rule is: the smaller the focal spot of the X-ray tubes used, the smaller the details which can normally be distinguished. On the other hand, the finite resolution of the detector can only be compensated by an increase of the geometrical magnification, insofar as the geometrical blurring determined by the focal spot is not a constraint. For a given source-detector-pair, there is a magnification factor to which both the blurring contributions are equivalent. The following

example should clarify this: when the focal size of an X-ray tube is  $f = 5 \mu\text{m}$  and a digital flat screen detector has  $p = 50 \mu\text{m}$  pixel distance, a magnification factor of  $V = 11$  ensures equivalence, i.e.,  $u_{\text{geo}} = p$ . As well an effective blurring of approximately  $4.5 \mu\text{m}$  results in the object, which is somewhat smaller than the size of the focal spot and considerably smaller than the detector's intrinsic resolution.

A decisive factor for the feasibility of X-ray imaging in industrial development and production is the use of digital detectors. They enable an immediate recording of the image as well as methods of subsequent image processing including an automatic evaluation. The use of modern flat panel detector-technology in conjunction with computer digital data processing have made it possible to build automatic radioscopy systems for nondestructive examination and in addition to this to develop further methods of three-dimensional computed tomography.

---

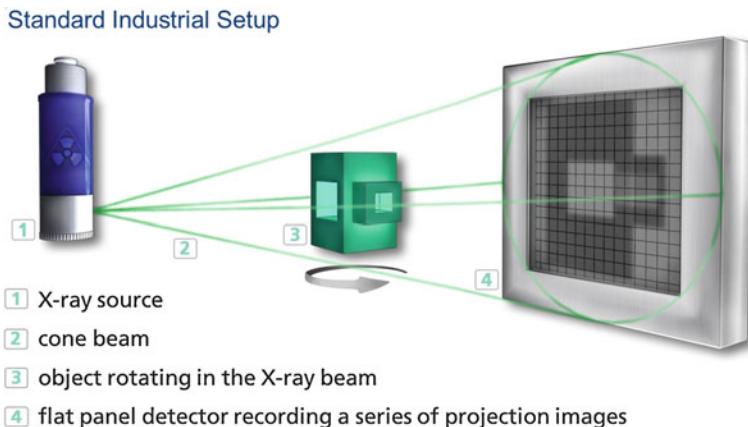
## Process for Sectional Imaging with X-Ray Radiation

In general, the term computed tomography (CT) can be understood to mean the computer-aided calculation of sectional images (Greek: *tomos* = section) for well-defined levels and orientations within a three-dimensional object (Hounsfield 1973; Buzug 2008; Natterer 1986; Katsevich 2004; Herman 2009; Feldkamp et al. 1984). The basis for all forms of CT is a series of projection images of the object being examined, each of which is measured as described in the previous paragraph. The resulting set of primary data is usually called the *sinogram*. Depending on the procedure of data acquisition, we can distinguish between several methods of CT. These all come along with their own specific method of image reconstruction, which means the calculation of the material characteristics inside a layer of the volume to be inspected.

Mathematically, strictly speaking, the reconstruction requires projection images from all directions of a half circle around the object (since in first order there is no difference in measuring an attenuation value along an X-ray in forward or backward direction).

In practice it does not matter whether the object is rotated inside the X-ray beam between a fixed source and an equally stationary detector or whether the source and radiation sensor turn synchronously around the stationary object. With this in mind, computer tomographic processes which are based on complete sets of data with radiation directions from  $180^\circ$  are sometimes grouped under the term axial CT.

As a rule, industrial CT-systems are carried out as real three-dimensional processes with cone-shaped radiation geometry. While medical CT-scanners move around a resting patient, with industrial systems the complete object is located in a fixed beam cone and undergoes the CT process in one single rotation (Fig. 4). In this process, as a large amount of measuring data is accrued during a very short time, highly efficient reconstruction algorithms must be available to avoid subsequent delays during data processing. With the present state of technology the duration of one rotation during which the complete object is scanned from projection angles over  $360^\circ$  can vary between 1 s and several hours, depending on the particular demands on spatial resolution and contrast-noise-ratio (Fuchs and Kalender 2003).



**Fig. 4** Principles of data recording in industrial cone beam-CT

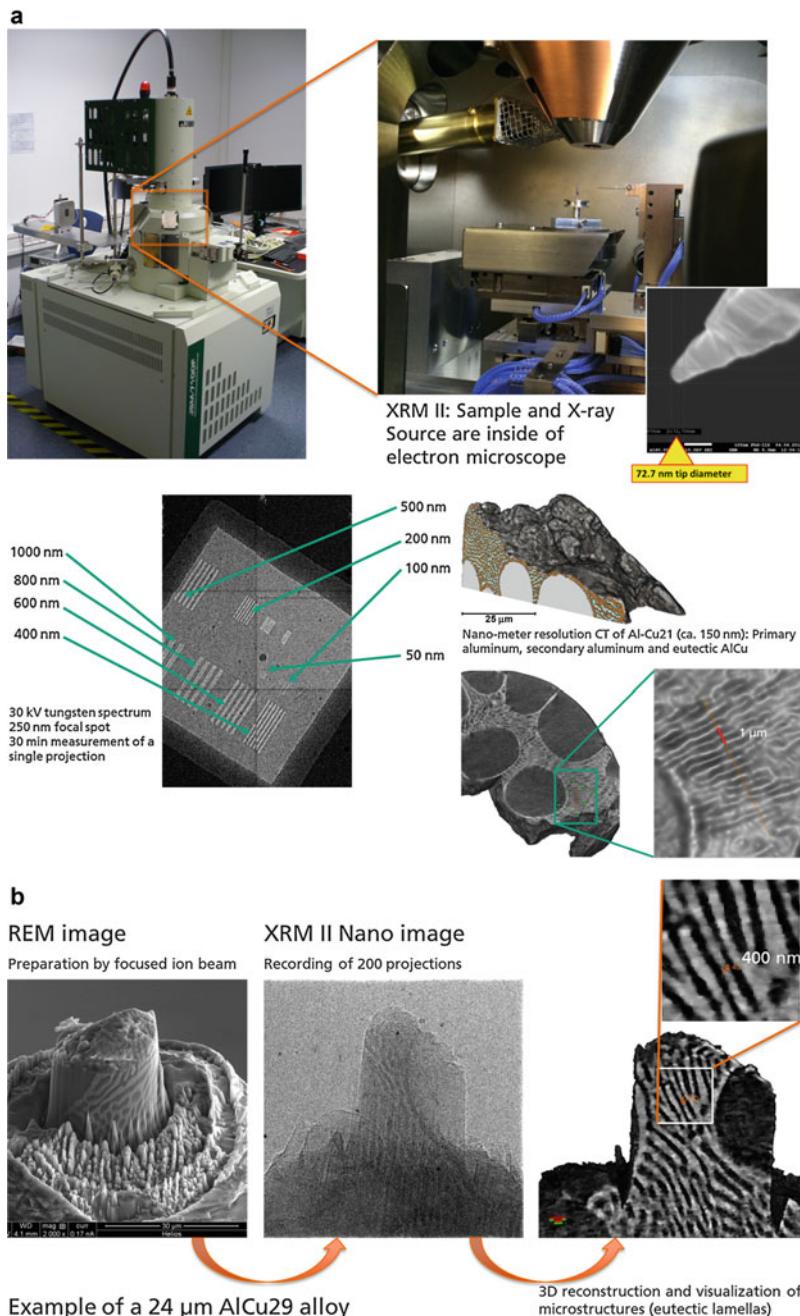
## Micro-Und Nano-CT

Micro-CT ( $\mu$ -CT), in particular, is a process which has no counterpart in the area of X-ray diagnostics for human medicine. In industry, it is used to measure small components with spatial resolutions in the  $\mu\text{m}$ -range and lower (recently also with 100 nm or even better). For this process, the use of an X-ray source with very low dimension of the X-ray focus is required. The size of focal spots  $f$  of industrial micro-focus-X-ray tubes ranges between 0.1 and 50  $\mu\text{m}$  (Salamon et al. 2008a; Zabler et al. 2012).

As small objects permit high geometrical magnifications, the negative effect of scattering radiation due to the cone beam geometry of the  $\mu$ -CT is reduced significantly. The reason for this is that when the geometric magnification is high, the object is positioned very close to the X-ray source, whereas the detector is located comparatively far away (standard magnifications are around  $V = 100$  and more). As the distance between object and detector increases, the spatial angle of the scatter profile seen from the detector decreases significantly. Figure 5 shows a newly developed Nano-CT system (Stahlhut et al. 2014, 2013; Mayo 2002) and first results of a Nano-CT-measurement using the example of an aluminum-alloy sample.

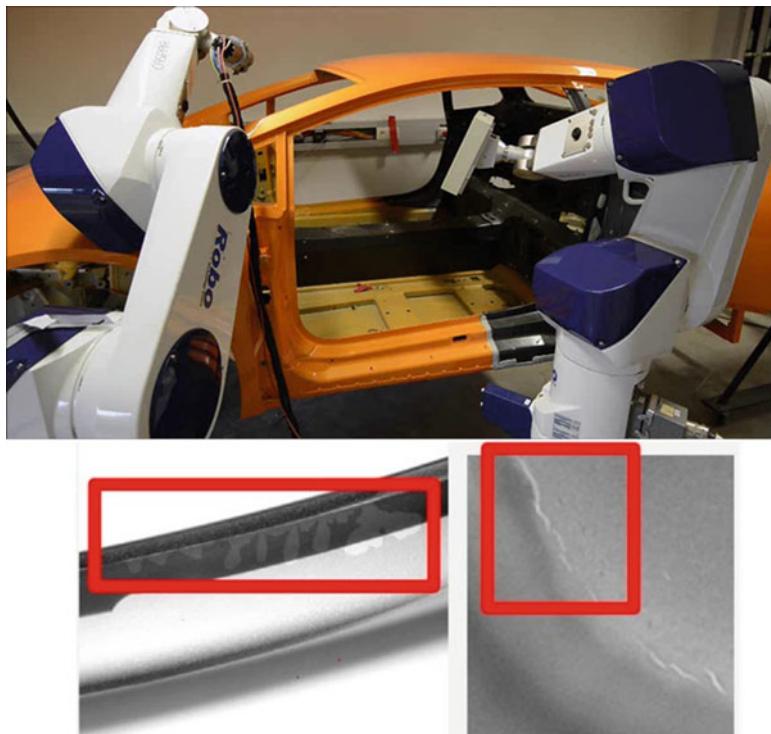
## New CT-Methods Based on Iterative Reconstruction Methods

One of the current, most interesting developments in the field of industrial CT is taking place in the area of image reconstruction. For a few years now, the processing power of commercially available personal computers, together with a high-performance graphic card (GPU), has been sufficient to calculate volume images with dimensions of the order of 1000 pixels cubed using iterative techniques within acceptable time.

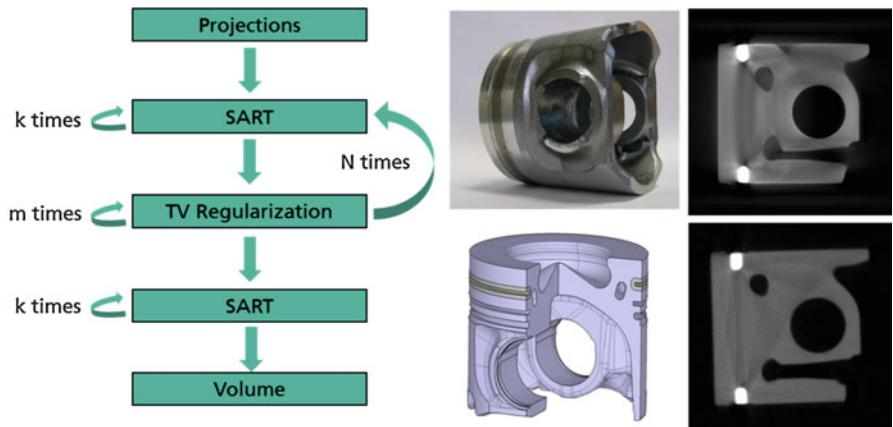


**Fig. 5** (a) Nano-CT-System XRM II at the Chair for X-Ray Microscopy LRM Julian-Maximilians-University Würzburg, Germany (Software: J. Engel, reconstruction K. Dremel and J. Dittmann, data measurement P. Stahlhut). (b) First results: on the left the projection image of a resolution test sample is shown. Right hand side: the 3D-reconstruction of a sample from an aluminum-alloy is presented with approximately 150 nm 3-D spatial resolution

These iterative techniques are based on a completely different approach to solving reconstruction problems, which consists of using a series of radiographic images (projections) to guess the original spatial distribution of the X-ray attenuation coefficient (the volume image of the object being examined). An iterative technique, also known as “Algebraic Reconstruction Technique” (ART) (Gordon et al. 1970), solves this problem, by successively adapting the volume images to the measured values in a series of iterations, so that the virtual projections generated in the computer (with a relatively high numerical effort) come as close as possible to those actually measured. The great advantage of ART is that it makes no intrinsic demands on the number or the directions of the radiation of projection images. Therefore, ART is superior to all conventional, analytical techniques when the projections are either irregularly distributed in the space, incomplete, or when only a small number of them can be determined.



**Fig. 6** Robot-based CT in the production process, as shown in the photograph above, is today just about being established. Two robots, one of them (left hand side) carrying the tube and the other a flat panel detector (the right one) constitute a highly flexible 3D X-ray system. This can be used, e.g., for high-resolution inspection of body shell segments. The reconstructed volume data allow a thorough inspection of join connections as for instance adhesive bonds (lower left image), welding seams (lower right), soldered joints, etc. as well as discontinuities in CFRP components. Thereby the main challenge is the complex calibration for correcting robot inaccuracies

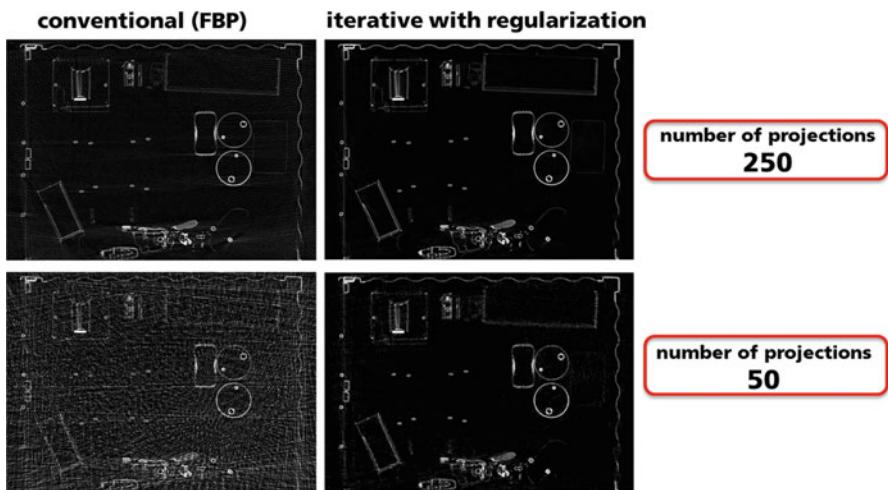


**Fig. 7** Thanks to the use of innovative reconstruction methods (schematic diagram: left), the number of projections required for high-quality sectional images can be reduced significantly. In the case of process-integrated examination of motor pistons, the direct result is a reduction of the measuring times from 20 min down to 30 s, i.e., by a factor of 40. Middle: a photo of the piston and corresponding 3D-image. Top right: cross-section conventionally reconstructed, bottom right: slice image generated by an iterative algorithm including a regularization step

In the context of industrial practice, robot-aided CT (cf. Fig. 6) can profit from algebraic reconstruction methods, as these allow virtually any deviation from the conventional circular path of the source around the object. In addition to this, iterative approaches can also be combined with so-called regularization methods, which originate from the general theory of compressive sensing (Zou and Pan 2004; Dittmann 2009; De Man and Fessler 2009; Sidky and Pan 2008). By using this, it is possible to greatly reduce the number of projections to be measured. This automatically leads to a significant reduction in measuring times and thus makes it possible to integrate and use a CT in a production line. The first realizations of such highly complex implementations of regularizing, iterative methods are the monitoring of motor pistons at production speed (cf. Fig. 7) and the rapid inspection (i.e., within in a few minutes) of sea freight-containers for illegal freight or freight which could be a public safety hazard (e.g., cigarettes or weapons of mass destruction, respectively) (Fig. 8).

## Outlook on the Future Development of Industrial X-Ray Imaging

As in the last few years, the innovative impulses for industrial imaging have originated in the progressive technologies of data- and information processing, it is to be expected that in the coming years, processes from the field of artificial-intelligence research will open up completely new aspects for X-ray-CT. An example of this is the possibility in the near future of the autonomous, i.e., independent and intelligent self-optimization of an X-ray inspection system (key word: machine



**Fig. 8** Comparison of a conventional reconstruction algorithm with the result of an iterative method combined with a regularization step (“Total Variation Minimization” TVM). Here we can see a sectional image of an actual sea freight container, left with a conventional method (the “Filtered Back projection” FBP) and right with the new complex iterative technique. Even when the number of projections used is reduced by a factor of 5, the latter technique still delivers a virtually identical image quality, whereas the quality of the conventional reconstruction deteriorates rapidly. Thus, the decrease in the number of measured projections required can in turn lead to a drastic reduction of the inspection time required in freight transport

learning). Autonomous sensor networks will be created in conjunction with sensor systems. These networks, in conjunction with virtual twins of the sensor systems, will support and maintain industrial products in an economically optimum condition during the complete product lifetime cycle.

## References

- Buzug TM (2008) Computed tomography: from photon statistics to modern cone-beam CT. Springer, Berlin. ISBN-13: 978-3540394075
- De Man B, Fessler JA (2009) Statistical iterative reconstruction for X-ray computed tomography. In: Censor Y, Jiang M, Wang G (eds) Biomedical mathematics: promising directions in imaging, therapy planning, and inverse problems. Medical Physics Publishing, Madison
- Dittmann J (2009) Tomographic reconstruction from few projections based on the theory of compressed sensing. Master thesis, Chair for X-ray microscopy, Julius-Maximilians-University, Würzburg
- Feldkamp LA, Davis LC, Kress JW (1984) Practical cone-beam algorithm. J Opt Soc Am 6:612
- Fuchs T, Hanke R (2008) Task-driven design of X-ray systems for industrial inspection. In: IEEE nuclear science symposium conference record. <https://doi.org/10.1109/NSSMIC.2008.4775230>
- Fuchs T, Kalender W (2003) On the correlation of pixel noise, spatial resolution and dose in computed tomography: theoretical prediction and verification by simulation and measurement. Phys Med XIX(2):153–164

- Gordon R, Bender R, Herman GT (1970) Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *J Theor Biol* 29:471–481
- Hemberg O, Otendal M, Hertz HM (2003) Liquid-metal-jet anode electron-impact X-ray source. *Appl Phys Lett* 83:1483–1485
- Herman GT (2009) Fundamentals of computerized tomography: image reconstruction from projections, 2nd edn. Springer, Dordrecht. ISBN 978-1-85233-617-2
- Hounsfield GN (1973) Computerized transverse axial scanning (tomography): part 1. Description of system. *Br J Radiol* 46:1016–1022
- Hubbell JH (1982) Photon mass attenuation coefficients and energy-absorption coefficients from 1 keV to 20 MeV. *Int J Appl Radiat Isot* 33:1260–1290
- Hubbell JH, Seltzer SM (1989) Tables of X-Ray mass attenuation coefficients and mass energy-absorption coefficients. [Online] National Institute of Standards and Technology. <http://www.nist.gov/pml/data/xraycoef/>
- Katsevich A (2004) Improved exact filtered back-projection algorithm for spiral CT. *Adv Appl Math* 32:681–697
- Mayo SC (2002) Quantitative X-ray projection microscopy: phase-contrast and multi-spectral imaging. *J Microsc* 207:79–96
- Natterer F (1986) The mathematics of computerized tomography. B.G. Teubner, Stuttgart. ISBN 0-471-90959-9
- Otendal M (2006) A compact high-brightness liquid-metal-jet X-ray source. Doctoral thesis, Department of Applied Physics, Royal Institute of Technology, Stockholm
- Salamon M, Hanke R, Krüger P, Sukowski F, Uhlmann N, Voland V (2008a) Comparison of different methods for determining the size of a focal spot of microfocus X-ray tubes. *Nucl Inst Methods Phys Res A* 591:54–58
- Salamon M, Hanke R, Krüger P, Uhlmann N, Voland V (2008b) Realization of a computed tomography setup to achieve resolutions below 1 μm. *Nucl Inst Methods Phys Res A* 591:50–53
- Salamon M, Burtzlaff S, Voland V, Sukowski F, Uhlmann N (2009) Upcoming challenges in high resolution CT below 1 micron. *Nucl Instrum Methods Phys Res A* 607:176–178
- Scholz O, Schmitt P, Kube M, Behrendt R, Uhlmann N (2009) Improvements in detector design for X-ray inspection of cast parts. *SAE Int J Mater Manufac* 2:134–139
- Sidky EY, Pan X (2008) Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Phys Med Biol* 53:4777–4807
- Stahlhut P, Ebensperger T, Zabler S, Hanke R (2013) Laboratory X-ray microscopy using a reflection target system and geometric magnification. *J Phys Conf Ser* 46:1–4
- Stahlhut P, Ebensperger T, Zabler S, Hanke R (2014) A laboratory X-ray microscopy setup using a field emission electron source and micro-structured reflection targets. *Nucl Instrum Methods Phys Res, Sect B* 324:4–10
- Sukowski F, Yaneu JF, Salamon M, Ebert S, Uhlmann N (2009) Virtual detector characterization with Monte-Carlo-simulations. *Nucl Instrum Methods Phys Res, Sect A* 607:253–255
- Zabler S, Fella C, Dietrich A (2012) High-resolution and high-speed CT in industry and research. In: SPIE conference: developments in X-ray tomography VIII, vol 8506
- Zou Y, Pan X (2004) Exact image reconstruction on PI-lines from minimum data in helical cone-beam CT. *Phys Med Biol* 49:941–959



# Grazing Incidence X-Ray Reflectivity and Scattering

33

Brian K. Tanner

## Contents

Introduction and Historical Background .....	1182
Basic Physics Underpinning the Method .....	1183
Specular Scatter from an Ideal Sample .....	1184
Specular Scattering from a Rough or Near-Surface Graded Material .....	1190
Diffuse Scatter from Rough Surfaces and Interfaces .....	1192
Metrology of Thin Single Films Using Specular Reflectivity .....	1197
Metrology of Multiple Films Using Specular Reflectivity .....	1201
Metrological Analysis of Reflectivity Profiles: How to Extract Material Parameters from the Data .....	1202
NDE Using X-Ray Reflectivity in the Laboratory .....	1205
NDE in the Fab .....	1207
Conclusions .....	1211
References .....	1211

## Abstract

The nondestructive evaluation (NDE) technique of grazing incidence X-ray reflectivity is described and set in its historical context. Use of X-ray reflectivity to determine near-surface density is described and the determination of surface interface width presented. It is shown how measurement of the diffuse scatter is necessary in order to distinguish topological surface roughness from compositional grading, as both effects have an identical effect on the true specular scatter. Metrology of single and multiple films is presented, and it is shown how automated fitting of model structures enables the user to extract film thickness with a precision of typically 20 pm. It is demonstrated that the limits of film

---

B. K. Tanner (✉)

Department of Physics, Durham University, Durham, UK  
e-mail: [b.k.tanner@durham.ac.uk](mailto:b.k.tanner@durham.ac.uk); [b.k.tanner@dur.ac.uk](mailto:b.k.tanner@dur.ac.uk)

thickness suitable for such metrology are between 2 nm and 1  $\mu\text{m}$  on flat and smooth substrates. In the context of use of X-ray reflectivity as an NDE technique in a manufacturing environment, there is discussion of approaches to very rapid data collection.

---

## Introduction and Historical Background

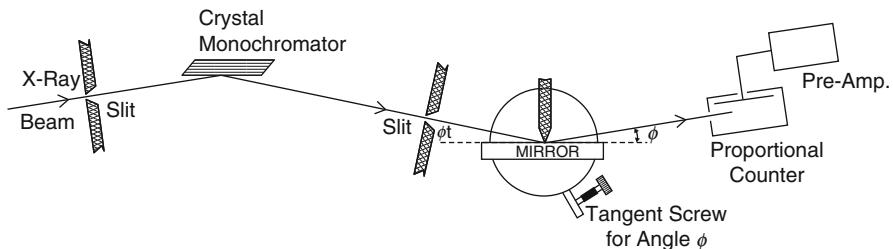
X-rays of energy about 10 keV, that is, of wavelength about 0.15 nm, have been found to have a refractive index a little below unity. This carries the intriguing implication that the phase velocity of X-rays exceeds the velocity of light in a vacuum. While students of physics attempt to reconcile this observation with what they thought they understood about the limits of transmission of information in the universe, the more practically focused of them recognize that, at some angles, total external reflection will occur when X-rays impinge on the surface of materials.

In 1930, Heinz Kiessig was studying the angular range of total external reflection of X-rays from various surfaces by rotating nickel-coated glass plates with respect to the direction of the incident X-ray beam. The reflected intensity for each angle was recorded on a photographic plate. For thick films, he observed a rapid fall-off in intensity beyond a critical angle, but for thin nickel films, he observed an increase and subsequent fall again of the intensity beyond the critical angle. Using an ionization chamber as a detector, he observed two fringes which he correctly interpreted as arising from interference of waves specularly reflected at the nickel-air and the glass-nickel interfaces (Kiessig 1930). From the angular fringe spacing, he deduced that the film thickness was 0.22 nm. During the year following this pioneering observation, Kiessig published two more papers, giving details of the instrumentation and further observations of total external reflection (Kiessig 1931a) and the interference from thin films (Kiessig 1931b).

Surprisingly, the technique was then somewhat neglected, and except for one pioneering group at Cornell University in the United States, led by L.G. Parratt, there was little work until the 1980s, when modern thin-film technologies started to blossom. Parratt's 1954 work, as we will see later, has been crucial to the development of the technique as an industrial nondestructive evaluation method. Although principally concerned with measuring the density of thin films, Parratt developed both a reflectometer with the characteristics of most current instruments (Fig. 1) and a recursive theory of the reflectivity from multiple layer structures that form the basis of current modeling approaches to the interpretation of X-ray data (Parratt 1954).

The Cornell group produced two more papers in the late 1950s studying the density of thin films of Cu (Wainfan et al. 1959) and the effect of oxidation on films of Cu, Ni, Ge, and Se (Wainfan and Parratt 1960). In the first of these papers, Wainfan et al. (1959) use the interference fringes as a metrological tool for determination of film thickness and address for the first time the comparison of numerical values with those obtained by optical interferometry.

It was not until 1977 that further thin-film applications were reported, but neither Isherwood (1977) nor, 5 years later, Martinbouyer and Dethy (1982) published their work in widely read mainstream journals. As a result, it was not until the late 1980s



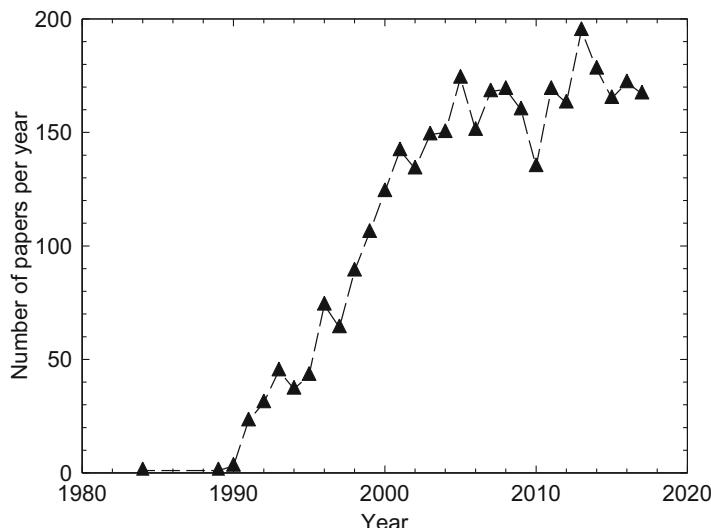
**Fig. 1** Original diagram (Fig. 1) from Parratt (1954) showing the geometry of the grazing incidence X-ray reflection apparatus. It is remarkably similar to the principles behind contemporary equipment. (Reproduced by permission of the American Physical Society)

that the technique started to be exploited widely and, interestingly, this was also the time when multi-layer X-ray reflection optics started to be used in the soft X-ray region (Meekins et al. 1986). At that time, reflective optics were becoming particularly important in the context of X-ray astronomy (Stevens et al. 1986). In 1987, Allain et al. (1987), reported grazing incidence reflectivity XRR studies of Langmuir-Blodgett films, which are effectively molecular multilayers, and this was quickly followed by a study of Langmuir-Blodgett films on water (Bosio et al. 1987).

The first applications of XRR to the metrology of semiconductor films were reported in 1987/1988, both of these being to determine the thickness of compound semiconductor epilayers from the spacing of the interference fringes in the specular scatter, observed as the angle of incidence is increased (Woronick et al. 1987; Lucas et al. 1988). This was rapidly followed by studies of carbon films (Toney and Brennan 1989), metal films (Chen and Heald 1989), and correlated roughness in multilayer films (Rosen et al. 1988). The first analyses of SiGe superlattices appeared at the beginning of the 1990s (Hudson et al. 1992; Baribeau 1993) and these years also saw the first application to magnetic multilayers, so-called spintronic systems (Huang et al. 1992; Hudson and Tanner 1992; Huang et al. 1993). From this point onward, there was a rapid linear growth in the use of the technique until about 2005, when there was a leveling off of activity (Fig. 2). The number of citations of the technique has, however, continued to rise linearly (Fig. 3), and X-ray reflectivity has now become a standard laboratory-based technique for thin film metrology. An extensive review was published some years ago by Chason and Mayer (1997) and a detailed discussion of work on semiconductors has been given in the book by Bowen and Tanner (2006). Application of X-ray reflectivity in a production environment is, however, still at a quite early stage.

## Basic Physics Underpinning the Method

The key feature behind all interferometry techniques is the difference in scattering power for the probing radiation on crossing an interface between one material and another. In the case of X-rays, this is primarily associated with the electron density



**Fig. 2** Number, as a function of year, of published papers cited in *Web of Science* containing keywords “X-ray reflectivity AND thin film”

and thus dependent on the atomic number of the material under investigation. Interference contrast between layers with substantially different atomic number is greater than those adjacent in the periodic table. In some case, for example, the magnetically important case of Cu and Ni<sub>80</sub>Fe<sub>20</sub> (permalloy), the electron density is almost equal and the contrast is extremely low.

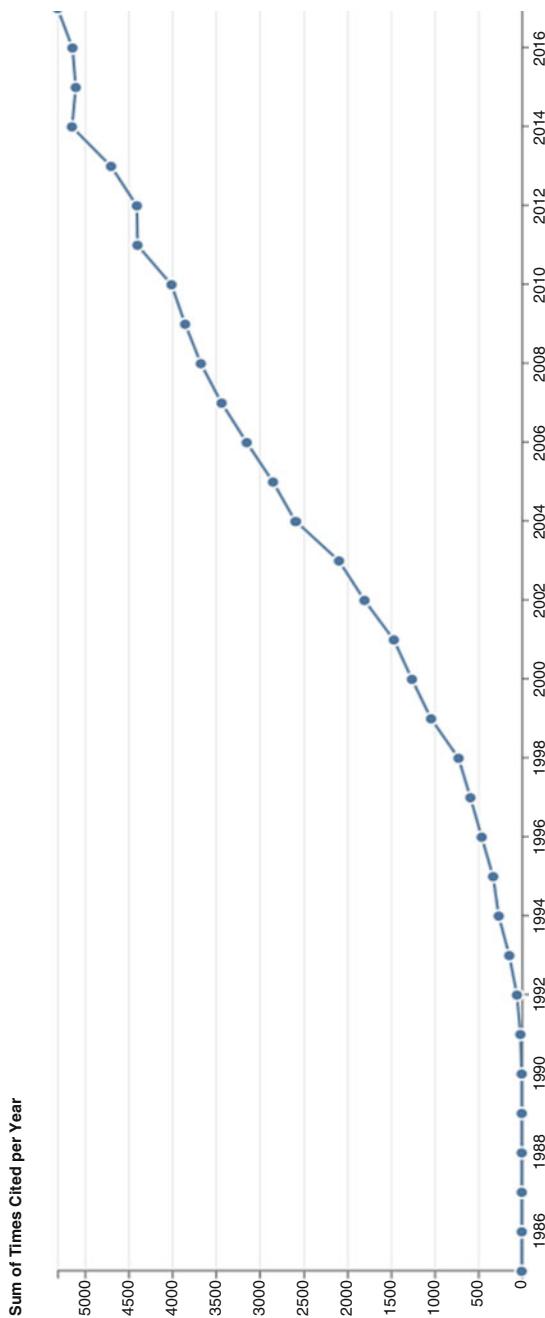
One of the reasons that make X-ray reflectivity less obvious to observe than optical interference effects is that the refractive index for X-rays in the 10 keV region is only a little less than unity. For NDE, this provides a problem because it implies that the scattering of X-rays by solids is very weak and at high angles, the reflected intensity is extremely small. Only at glancing incidence angle, close to the critical angle for total external reflection, can interference due to reflection at sequential interfaces be observed.

The principles of the technique can be understood by considering the reflection of any form of electromagnetic wave at a boundary, and the necessary mathematics is contained within most undergraduate textbooks on electricity and magnetism.

### Specular Scatter from an Ideal Sample

The principal requirement in standard continuum electromagnetic theory is that the amplitude and gradient of the X-ray wave are continuous across the interface between the vacuum (or to a very good approximation, air) and the solid.

As for visible light optics, Snell's law follows with



**Fig. 3** Number of citations per annum in *Web of Science* containing keywords "X-ray reflectivity AND thin film"

$$\cos \omega_I = n \cos \omega_T \quad (1)$$

where  $\omega_I$  is the incidence angle with respect to the surface and  $\omega_T$  is the refracted angle (Fig. 4). We see immediately that the critical angle for total external reflection, where  $\omega_T = 0$ , is given by

$$\cos \omega_c = n \approx 1 - \frac{\omega_c^2}{2} \quad (2)$$

The refractive index can be defined as a complex number  $n$  in order to additionally include the effect of absorption. Expressed in terms of scattering and absorption parameters  $\delta$  and  $\beta$  as

$$n = 1 - \delta - i\beta, \quad (3)$$

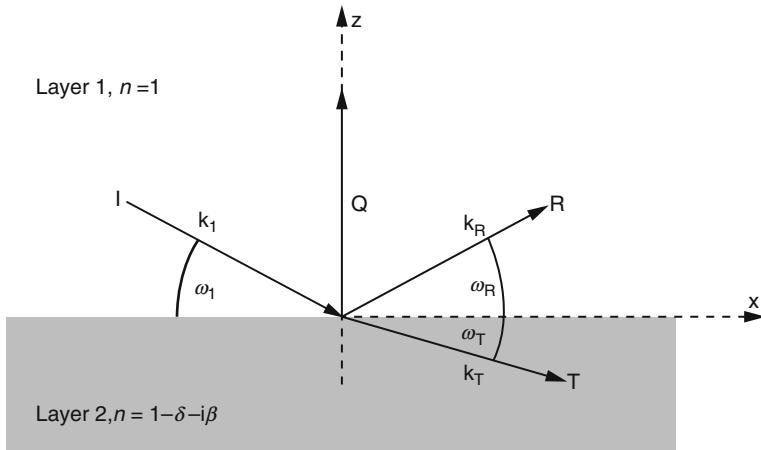
for the case of zero absorption ( $\beta = 0$ ), we have

$$\omega_c = \sqrt{2\delta}. \quad (4)$$

Below an incidence angle of  $\omega_c$ , the reflectivity is unity, and we have total **external** reflection. This is exactly analogous to the total internal reflection found in visible light optics.

The refractive index  $n$  can be derived from classical X-ray scattering theory and is given by

$$n = 1 - \frac{N_A}{2\pi} r_0 \lambda^2 \sum_j \frac{\rho_j}{M_j} f_j \quad (5)$$



**Fig. 4** Geometry for specular reflectance

where  $\rho_j$  is the density,  $f_j$  is the scattering factor of element  $j$  of mass number  $M_j$ ,  $N_A$  is Avogadro's number,  $\lambda$  is the X-ray wavelength, and  $r_0$  is the classical electron radius given by

$$r_0 = \frac{e^2}{4\pi\epsilon_0 m_e c^2} \quad (6)$$

where  $m_e$  is the electron mass,  $e$  the electron charge,  $c$  the velocity of light, and  $\epsilon_0$  the permittivity of free space.

The scattering factor  $f_j$  is a measure of the strength of the X-ray scattering for a particular element  $j$  and for small angles is given by

$$f_j = Z_j - f'_j - if''_j \quad (7)$$

where  $Z_j$  is the atomic number of element  $j$  and  $f'_j$  and  $f''_j$  are the real and imaginary parts of the dispersion correction.

Therefore, to first order, the scattering factor scales with atomic number, as stated earlier. It is worth noting at this point that there is a very great difference for the case of the equivalent technique of grazing incidence neutron reflectivity (Penfold 1991) which also became widely available in the early 1990s. Unlike the X-ray case, where the scattering is purely via the electromagnetic interaction, for the neutron case, the scattering has two contributions. There is firstly a magnetic dipole-dipole interaction between the magnetic moment of the neutron and the magnetic moment on the atom, which can be exploited using polarized neutron beams to probe the magnetization of thin films of spintronic materials (Blundell and Bland 1992). Polarized neutron reflectivity can be performed at both pulsed and continuous beam neutron scattering facilities, for example at ISIS at the Rutherford Appleton Laboratory in Oxfordshire, UK (Satchell et al. 2017) and the Institute Laue-Langevin in Grenoble, France (Cubitt and Fragneto 2002). The second interaction, which is observable for all materials, is via the strong interaction and the scattering is from the atomic nuclei, not the electron cloud. As the structure of the nucleus does not scale with the atomic number, there is a somewhat erratic variation in the scattering factor (or scattering length as used by the neutron community) through the periodic table. Thus, elements adjacent in the periodic table may have hugely different scattering lengths, permitting, for example, films of Zn and Cu to be distinguished clearly. Further, the scattering from hydrogen and deuterium is strong, unlike the X-ray case, and isotopic substitution of deuterium for hydrogen results in strong contrast. There has been much work on organic and soft condensed matter systems (Penfold 2002), with diblock copolymer systems being particularly amenable to study. Examples include Anastasiadis et al. (1989), Lewicky et al. (1998), Zhang and Somasundaran (2006), and Koenig et al. (1996). The X-ray and neutron techniques are thus complementary.

The X-ray critical angle  $\omega_c$  can be expressed as

$$\omega_c = \lambda \sqrt{\frac{N_A r_0}{\pi} \sum_j \frac{\rho_j}{M_j} (Z_j - f'_j)} \quad (8)$$

or

$$\omega_c = \lambda \sqrt{\frac{\rho N_A r_0 [X_1(Z_1 - f'_1) + X_2(Z_2 - f'_2) + \dots]}{\pi[X_1 M_1 + X_2 M_2 + \dots]}} \quad (9)$$

where  $\rho$  is the overall density and  $X_j$  is the atomic mass fraction of element  $j$  in the material.

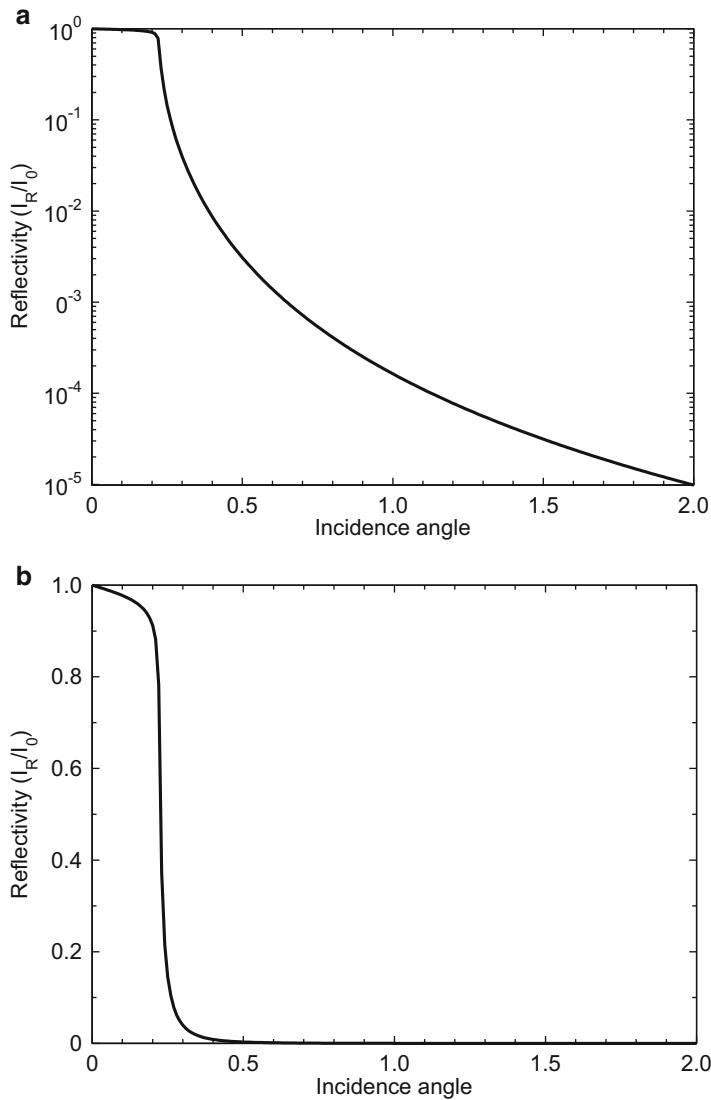
As the X-ray scattering factors are tabulated, it can be seen from Eqs. 8 and 9 that the near surface density can thus be measured directly from the critical angle. This was the focus of the early work of Parratt and colleagues cited earlier. As described below, the near-surface density measurement can be made simply from the position of the rapid fall in the specular reflectivity. The method was used extremely successfully to measure the  $sp^2$  to  $sp^3$  fraction in diamond-like carbon films by Ferrari et al. (2000).

Below the critical angle, the X-ray wave enters the material only as a decaying evanescent wave, the depth of penetration being typically only a few nanometres and whose value is dependent on the composition of the material. The reflectivity  $R$ , defined by the ratio of the reflected intensity  $I_R$  to the incident intensity  $I_0$ , is unity. Above the critical angle, the X-ray wave penetrates the material as a traveling wave and the reflected amplitude falls rapidly (Fig. 5). Application of the continuity conditions to the incident, transmitted and reflected wave amplitudes at a perfect interface yields the well-known Fresnel equations and these describe the X-ray scatter. Expanding to second order for small angles gives the reflectivity (i.e., the ratio of reflected to incident intensities), which varies as the inverse fourth power of the incidence angle. That is,

$$I \propto \omega_I^{-4} \quad (10)$$

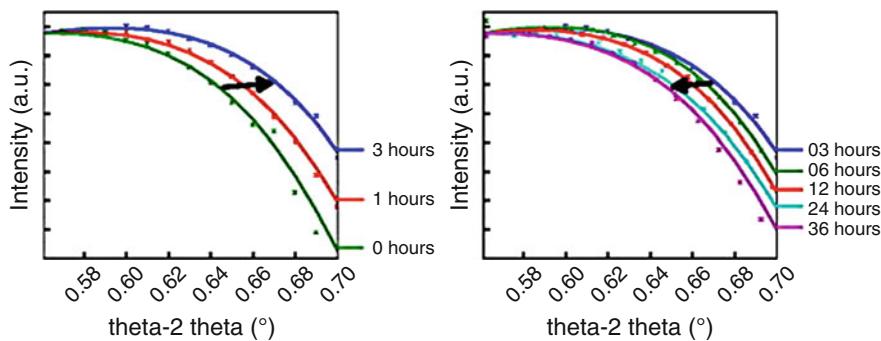
The above dependence is shown in Fig. 5 on both log and linear scales. Whereas the conventional log scale (Fig. 5a) shows the inverse fourth power fall beautifully, it is difficult to identify precisely the critical angle. The linear scale (Fig. 5b) provides a much clearer identification and the  $\omega_c$  value can be taken as the halfway point down the reflectivity curve.

Reference has already been made to the application of critical angle measurement for determination of the density and hence  $sp^3$  fraction of diamond-like carbon films (Ferrari et al. 2000; LiBassi et al. 2000). Other examples include the density of liquid Ga (Regan et al. 1995), of ceramic films (Veldhuis et al. 2014), and the porosity of low-k dielectric films for ULSI applications (Soles et al. 2003). An excellent recent example of the application to the measurement of changes in density, as a function of annealing time, of thin sputtered amorphous InGaZnO (a-IGZO) films on silicon has been published by Troughton et al. (2017). As seen in Fig. 6, as a function of time,



**Fig. 5** Calculated reflectivity of a perfectly smooth silicon substrate on (a) log scale, (b) linear scale

the critical angle changes. At an annealing temperature of 200 °C, for the first 3 h of annealing, the critical angle increased, indicating that the density of the film had increased. Longer annealing times subsequently led to a decrease in the critical angle, that is, the film density subsequently decreased. For 300 °C annealing, there was no densification.

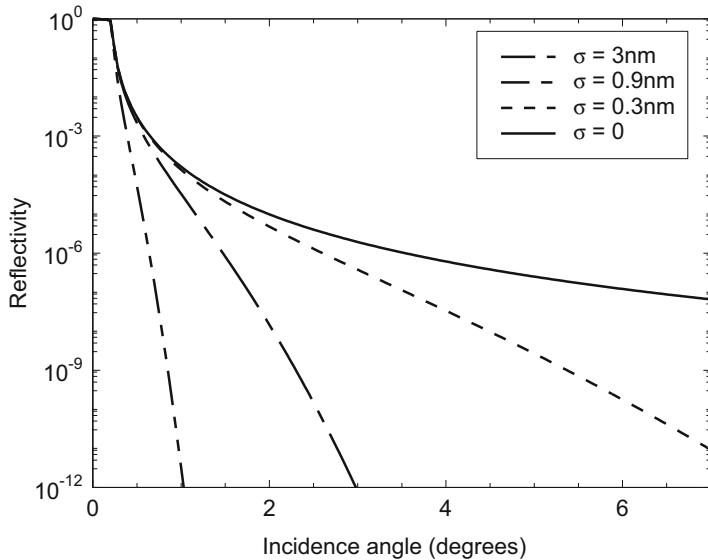


**Fig. 6** Region of the specular reflectivity from an a-IGZO film as a function of annealing time at 200 °C. (a) 0–3 h, (b) 3–26 h. The solid curves are best fits to the data. In (a) there is an increase in the critical angle with time, whereas in (b) the critical angle decreases with subsequent annealing. Plots are linear. (From Troughton et al. 2017, reproduced by permission of the American Institute of Physics)

### Specular Scattering from a Rough or Near-Surface Graded Material

We should note, at this point, that to measure the specular scatter, i.e., the mirror-like reflectance, we must scan the detector angle at the same rate as the incidence angle, so that the angles of incidence and reflection are always equal. We have also noted above, in Eq. 10, that for a perfectly abrupt surface, the specular reflectivity falls off as the inverse fourth power of the incidence angle. This immediately prompts the question; on what length scale can we define the surface to be ideally abrupt? Firstly, we note that as the scattering angles and hence scattering vectors are small, the structure on an atomic scale is not probed. To study this, we must resort to high-angle scattering, namely, X-ray diffraction. Materials can be treated as continuous when examined by X-ray reflectivity. Secondly, we must recognize that surfaces are intrinsically rough on various length scales. Such rough surfaces scatter X-rays out of the specular direction and thus the intensity, i.e., the reflectivity falls below the ideal value. The rate of this increase rate of fall-off depends on the roughness. The higher the roughness, the greater the scattering out of the specular and the greater the rate of fall off. By measuring the rate at which the intensity falls with incidence angle (and of course reflectance angle) one can determine the roughness of the interface. Figure 7 gives examples of how the reflectivity falls as a function of angle for Si surfaces of increasing r.m.s. roughness  $\sigma$ .

It is evident that for roughness above about 1 nm, the fall as a function of angle is so rapid that it becomes difficult to fit data to the model. In any case, the theoretical basis for the calculation becomes suspect at such roughness levels. The first observation on the method as an NDE technique is that it is appropriate only for extremely smooth surfaces. Also, because the scattering is so sensitive to angle and the beam is



**Fig. 7** Calculated specular reflectivity of silicon as a function of incidence (and detector) angle for different values of r.m.s. roughness  $\sigma$

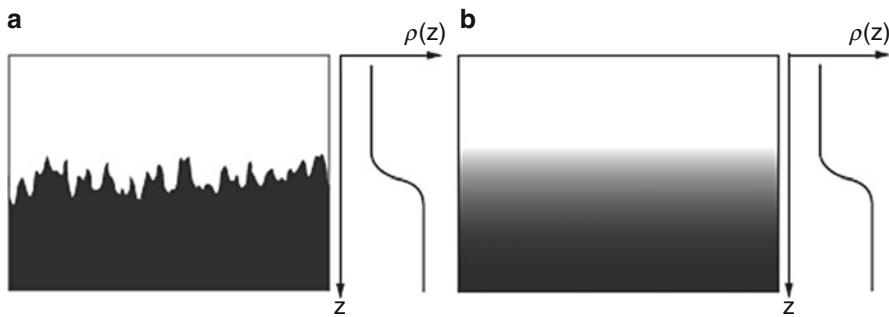
spread out across the surface by a factor of  $(\sin \omega_I)^{-1}$  the surface must also be very flat on a macroscopic scale.

The scattering vector  $\mathbf{Q}$ , is defined as the vector difference between the scattered wavevector  $\mathbf{k}_R$  and in incident wavevector  $\mathbf{k}_I$ , that is

$$\mathbf{Q} = \mathbf{k}_R - \mathbf{k}_I \quad (11)$$

Note that  $|\mathbf{k}_R| = |\mathbf{k}_I| = 1/\lambda$ ,  $\lambda$  being the X-ray wavelength. Inspection of Fig. 4, however, reveals that  $\mathbf{Q}$  is perpendicular to the surface and thus there can only be information about the surface in the direction normal to the surface. As we will see later, this is not a problem for the precise measurement of the thickness of films, but it does restrict information about the interface topology. We see from Fig. 8, which shows the density profile perpendicular to the surface averaged along the surface, that the averaged density profile of a topologically rough interface is identical to that of a surface in which the composition is graded with depth. Specular reflectivity therefore measures the interface width and does not distinguish between topological roughness and compositional grading. This is an extremely important point, as there are many careless references in the literature to roughness measurement, when in fact only the interface width has been determined.

The theoretical treatment of the effect of a rough surface on the specular reflectivity was published by de Boer (1994, 1996) and tends to be incorporated into most commercial and open source software for fitting X-ray reflectivity profiles to model structures. There are a number of sites that enable the calculation of the reflectivity



**Fig. 8** Averaged density profile as a function of coordinate  $z$  for (a) a topologically rough interface and (b) a compositionally graded interface. Note that these are indistinguishable if there is only information in the  $z$  direction

profile with inclusion of roughness or grading, and that of Henke ([http://henke.lbl.gov/optical\\_constants/](http://henke.lbl.gov/optical_constants/). Accessed July 2018) at the Center for X-Ray Optics at Berkeley Laboratory is particularly convenient.

The papers on the kinetics of surface roughening by You et al. (1993) and the kinetics of the growth of gallium oxide on gallium arsenide surfaces by Tanner et al. (2001) are examples of X-ray reflectivity studies of surface roughness and grading, respectively. Polishing of semiconductors is at a very advanced stage, and a topological roughness of 0.1 nm, as measured by atomic force or scanning tunneling microscopy, is not unusual for Si wafers from commercial vendors. As will be described below, measurement of topological roughness is best performed by measuring the diffuse scatter but measurement of compositional grading can be done by specular reflectivity.

## Diffuse Scatter from Rough Surfaces and Interfaces

As is the case for optical reflectivity, the misorientation of part of the surface results in X-rays being scattered out of the specular beam into a different direction. Clearly, the greater the roughness, the greater the scatter out of the specular beam and the more rapid will be the fall in the specular intensity. If, however, the detector is set away from the specular condition, the amount and distribution of this diffuse scatter can be measured directly. As any increased fall in specular reflectivity due to compositional grading is not associated with diffuse scatter being produced, the two components can be separated by measurement of both specular and diffuse scatter. In both cases, the specularly reflected intensity  $I_{\text{spec}}$  is reduced with respect to the value for a perfectly smooth surface  $I_{\text{ideal}}$  by a phase factor. For topological roughness, this is given by,

$$I_{\text{spec}} = I_{\text{ideal}} \exp[-Q_z^2 \sigma_t^2] \quad (12)$$

where  $\sigma_t$  is the r.m.s. roughness amplitude and  $Q_z$  is the component of the scattering vector normal to the interface.

As the intensity difference ( $I_{\text{ideal}} - I_{\text{spec}}$ ) must have been scattered into the diffuse scatter, the total diffuse scatter intensity  $I_{\text{diff}}$  is simply

$$I_{\text{diff}} = I_{\text{ideal}} - I_{\text{spec}} \quad (12)$$

Thus,

$$I_{\text{diff}} = I_{\text{spec}} \exp[Q_z^2 \sigma_t^2] - I_{\text{spec}} \quad (13)$$

or

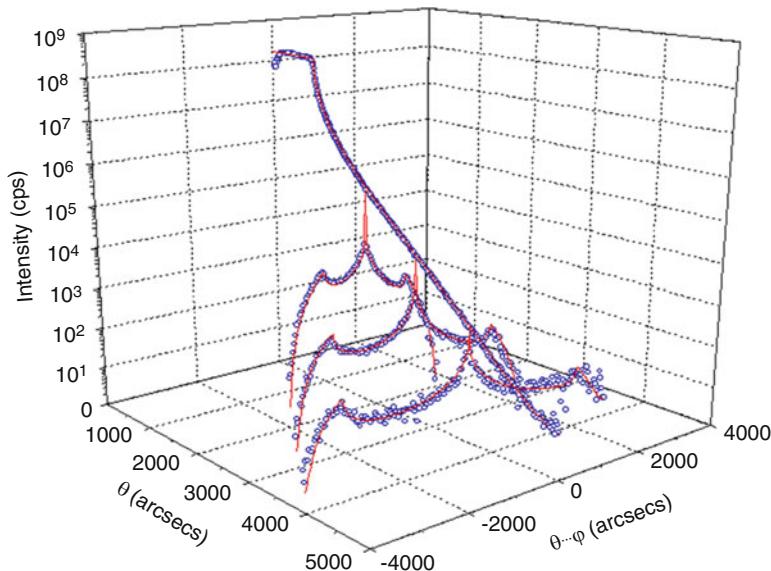
$$\sigma_t = \frac{1}{Q_z} \sqrt{\ln \left[ 1 + \frac{I_{\text{diff}}}{I_{\text{spec}}} \right]} \quad (14)$$

This so-called Born approximation works well only if the in-plane length scale of the roughness is quite large, as the diffuse scatter then lies close to the specular scatter. If the length scale of the roughness becomes small, the scatter is into a much larger angular region and the detector will collect only a fraction of the scatter. Equation 14 should be used with caution and preferentially interpreted qualitatively, an example being in the study of roughness of float glass (Pape et al. 1999).

A theory of the diffuse scatter, based on the distorted wave Born approximation, which takes into account the in-plane length scale by use of a fractal description of the interface, was developed by Sinha et al. (1988). It was extended to cover the case where there was also a compositional grade in the surface layers by Wormington et al. (1996). They also undertook a test of the theoretical predictions against measurements of the scattering from the surfaces of polished Zerodur® high stability oxide glass. The agreement between theory and experiment was excellent (Fig. 9) when a composition grading of the Zerodur® was included.

In these measurements, the diffuse scatter was measured by fixing the detector and scanning the sample direction with respect to the incident beam. This so-called rocking curve is the commonest form of data collection. For a perfectly flat surface, intensity would only be recorded when the angle of the incident beam with respect to the sample surface is half of the angle that the (fixed) detector makes with the (fixed) incident beam. The width of the peak is determined in this case by the beam divergence.

As seen in Fig. 9 and further exemplified in Fig. 10, the diffuse scatter is not uniform with respect to angle. When the roughness has a long in-plane length scale, the scatter is concentrated close to the specular peak and falls away, often with almost Lorentzian, or Lorentzian to the power  $2h$ , dependence (Fig. 10a). In this example, of the diffuse scatter from an almost 50 nm Au layer grown on a seed Fe layer on a (001) oriented MgO wafer, the best fit to the diffuse scatter is a Lorentzian to the power 0.7.



**Fig. 9** Specular and diffuse scatter from polished Zerodur®. Experimental specular data as a function of the angle  $\theta$  between the incident beam and the specimen are represented by the central curve. Measured rocking curves, for fixed detector angle  $2\theta$ , as a function of the displacement of the specimen from the specular position  $\theta - \phi$  are the other three curves. Solid lines are the theoretical fits to the data for a single set of material parameters (Wormington et al. 1996). (Reproduced by permission of Taylor and Francis Ltd., [www.tandfonline.com](http://www.tandfonline.com))

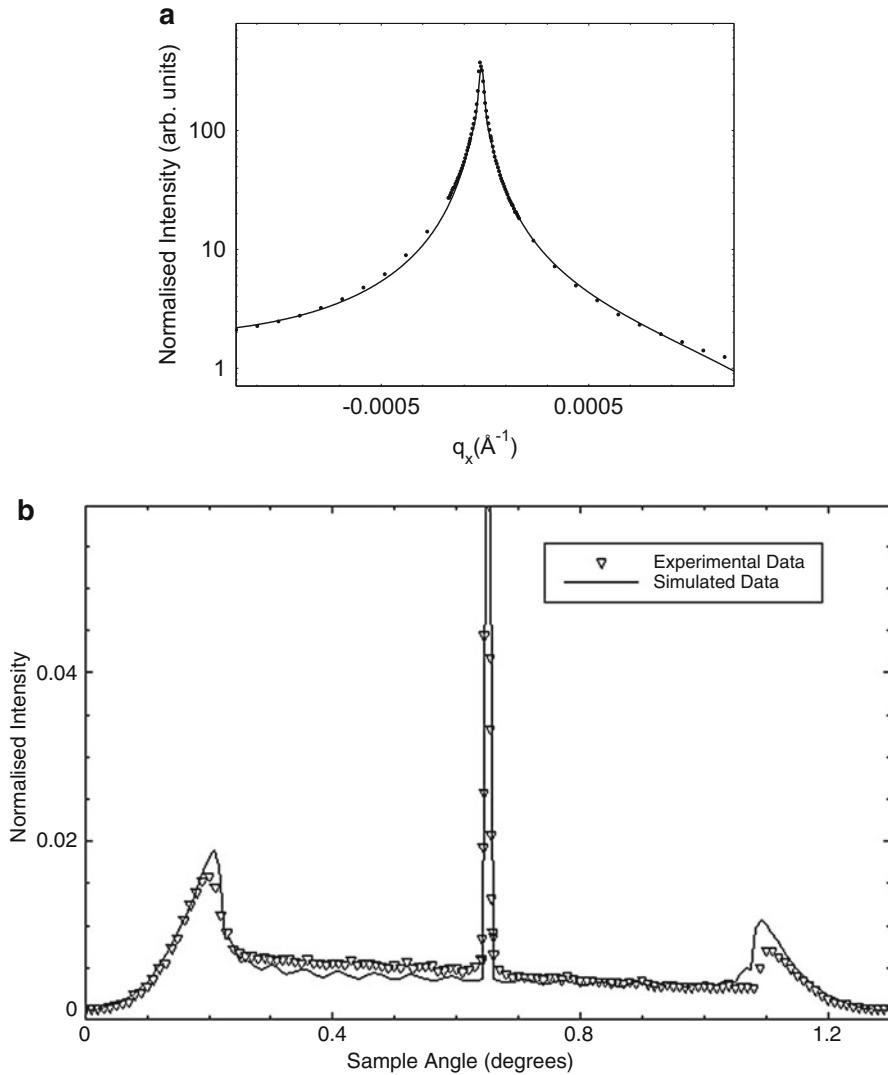
The fractal model Sinha et al. (1988) predicts the power law dependence observed in Fig. 10a for a surface with a fractal parameter of  $h = 0.35 \pm 0.05$ . Such a low fractal parameter is characteristic of initial island growth of Au (Cole et al. 2004). When the scatter is of this form, the in-plane correlation length  $\xi$  of the roughness can be estimated from the full width at half height maximum (FWHM) of the fitted curve. To a first approximation,

$$\xi = 1/\Delta Q_x \quad (15)$$

where  $\Delta Q_x$  is the FWHM in terms of the scattering vector component. Here,  $Q_x$  is related to the specimen angle, denoted variously as  $\omega_l$ ,  $\omega$  or  $\theta$ , detector angle  $\phi$  and X-ray wavelength  $\lambda$  by

$$Q_x = \frac{4\pi}{\lambda} \sin \frac{\phi}{2} \sin \omega \quad (16)$$

In the example given, the correlation length was found to be  $500 \pm 100$  nm, consistent with very large step lengths observed in the scanning tunneling microscope.



**Fig. 10** (a) Rocking curve of the scatter from an approximately 50 nm thick gold film grown on (001) MgO. The solid line is a fit to a Lorentzian function to the power 0.7. (b) Rocking curve of the surface of a single film of  $\text{Co}_{18}\text{Ag}_{82}$  alloy on a silicon substrate. The solid line is a theoretical fit using Wormington's model (Wormington et al. 1996) which is an extension of the theory of Sinha et al. (1988). A single in-plane correlation length  $100 \pm 2$  nm fits the data very well. ((a) Reproduced by permission of the Institute of Physics)

The example in Fig. 10b is of a surface that has a much smaller in-plane length scale (correlation length). Here the scatter is distributed much more evenly as a function of angle. In addition, we note the prominence of two additional peaks, one close to the position of grazing incidence and one close to the position of grazing exit from the surface.

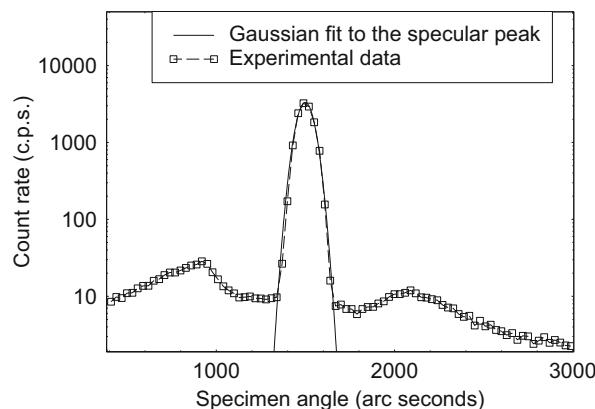
These two peaks were first noted by Yoneda in 1963 (Yoneda 1963) and have been named in his honor as “Yoneda wings.” They occur when the incidence angle or the exit angle equals the critical angle for total external reflection. Qualitatively, we can see why there is a rise in the diffuse scatter at these angles if we note that the low-angle wing occurs when the incident beam makes the critical angle, and at this point, the specularly reflected beam travels parallel to the surface. Thus, there is an increase in the diffuse scatter due to this longer path length in the material.

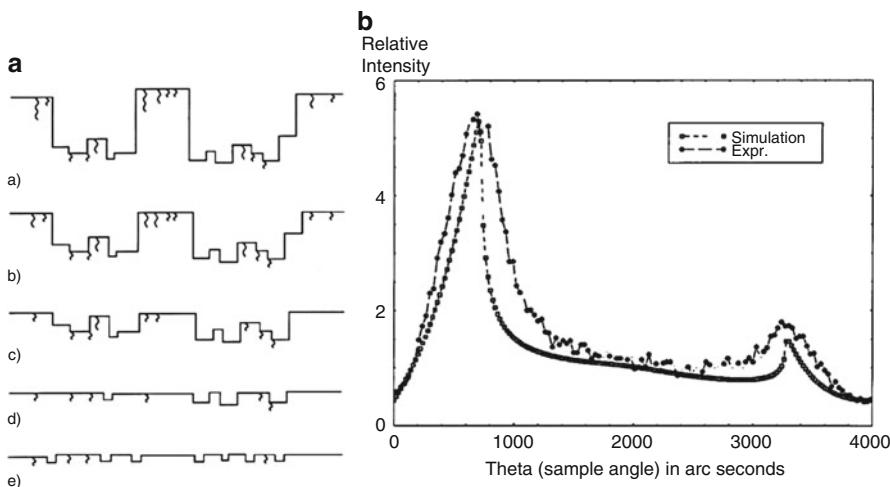
Quantitative measurement of the r.m.s. roughness using Eq. 14 is straightforward. The specular scatter peak in the rocking curve, the shape and width of which is determined by the beam geometry, should be fitted to a Gaussian function. Any data analysis software package can be used to measure the integrated intensity under this peak. The diffuse scatter can be determined from the total scatter integrated under the curve minus the specular scatter. An example is shown in Fig. 11.

The “Yoneda wings” provide a measure of the critical angle, which is determined by the surface density. Even when the roughness of the surface is too great for any true specular scatter to remain, the density can still be measured from the position of the Yoneda wings. Indeed, a very rough surface gives very strong diffuse scatter! An example is the study of the effect of polishing of alumina plates used for washers in mixer taps (faucets). The polishing pulls out alumina grains, (Pape et al. 2000), resulting in a reduction in the effective surface density of the material. The diffuse scatter comes primarily from the tops of the remaining grains (Fig. 12a). Because the projected coherence length of the X-ray wave is large, the change in effective density gives a shift in the position of the critical angle and hence in the position of the very prominent (Fig. 12b) Yoneda wings.

For semiconductors, it is often difficult to measure the diffuse scatter with laboratory-based equipment as the roughness achieved by chemi-mechanical polishing is so low that the resulting diffuse scatter is very weak. This is particularly true of silicon. (The data in the above examples were all taken at synchrotron radiation

**Fig. 11** Rocking curve for a GaAs sample that had been exposed to air for nearly 2 years. Detector angle of 3000 arc seconds ( $0.83^\circ$ ). The specular peak has been fitted to a Gaussian, enabling its component to be extracted and the true diffuse scatter can then be extracted from the integrated scatter under the total curve. Note that the plot is on a logarithmic scale



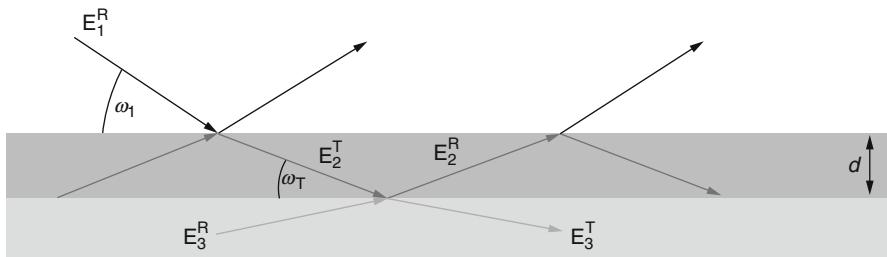


**Fig. 12** (a) Schematic diagram of the evolution of the effective surface density as a function of polishing, following grinding. As the polishing continues, the effective density increases. (b) An example rocking curve of a ground and polished alumina surface showing no true specular scatter but very clear Yoneda wings. (Pape et al. 2000). (Reproduced by permission of Taylor and Francis Ltd., [www.tandfonline.com](http://www.tandfonline.com))

sources.) Thus, for semiconductor thin film metrology purposes, one can often ignore the diffuse scatter. It is, however, good laboratory practice to measure the diffuse scatter in a scan of the specimen and detector in the ratio of 1:2 with the specimen displaced from the exact specular condition. This so-called off-specular  $\theta$ - $2\theta$  scan, with the specimen displaced typically  $0.1^\circ$  from the specular condition, measures the diffuse scatter approximately in the forward direction. This should then be subtracted from the specular measurement to obtain the true specular scatter. However, for semiconductor and spintronic systems, the correction is so small that it is often satisfactory to measure initially and not repeat the off-specular measurement for each individual sample.

## Metrology of Thin Single Films Using Specular Reflectivity

The key NDE application of X-ray reflectivity is in the metrology of thin films which are less than about  $1 \mu\text{m}$  in thickness. Above the critical angle, the X-ray wave penetrates into the sample and if there is a film on the substrate there will be reflected waves from both the top and bottom surfaces of the film (Fig. 13). Because the X-ray wave is coherent over this distance, there is interference between these two waves.



**Fig. 13** The path difference giving rise to the phase difference between waves reflected at the top and bottom surfaces of a thin film on a substrate

As the incidence angle changes, the path length between these reflected waves also changes. The associated change in the phase difference between these two waves results in interference being observed in the signal reaching the detector. Interference fringes are observed as the phase difference cycles through multiples of  $2\pi$  radians.

From Fig. 13, it is easy to see that the path difference between the waves reflected at top and bottom is  $2d \sin \omega_I$ . For constructive interference, we immediately have the equivalent of Bragg's Law, namely

$$n\lambda = 2d \sin \omega_I. \quad (17)$$

The period  $\Delta\omega_I$  of the interference, or Kiessig, fringes is

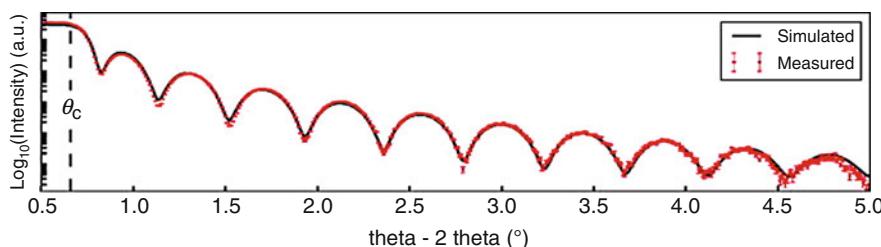
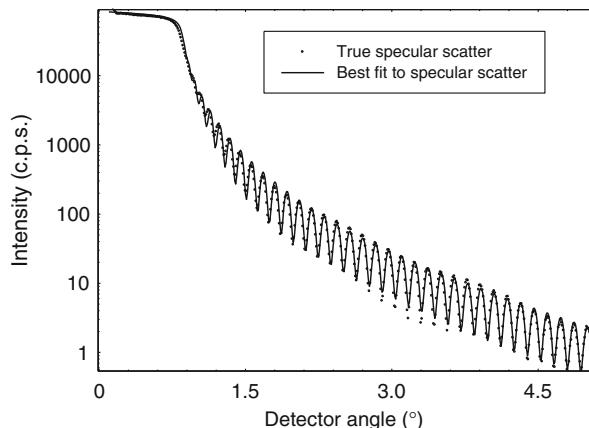
$$\Delta\omega_I = \lambda/2d \quad (18)$$

A single layer results in a single fringe period. An example (Fig. 14) of such a simple reflectivity profile is that from a single Au layer on an MgO substrate.

While the thickness of a single film can be determined just by measuring the interference fringe period, this is not recommended. Firstly, in the regions close to the critical angle, the scattering is strong and the simple picture needs modifying. There is often found a contraction of the fringe period in the low-angle region, an effect for which there is accurate compensation in the distorted wave Born approximation formalism. Thus, the normal procedure is to fit a model structure to the data and refine the model parameters until the difference between the experimental and simulated curves is minimized. This is what has been done in Fig. 14, resulting in a measurement of the film thickness of  $49.4 \pm 0.1$  nm with a top interface width, associated either with topological roughness or compositional grading of 0.1 nm and an interface width on the MgO of 0.5 nm.

Another beautiful example of a single period interference fringe pattern (Fig. 15) can be found in the paper of Troughton et al. (2017) which has already been referenced. There the single film of amorphous InGaZnO gave an almost textbook oscillatory fall in intensity. (Note that the sharp dips at the minima compared with the broad intensity maxima in the profile are entirely presentation artefacts of

**Fig. 14** True specular scatter from a single layer of Au grown by MBE on a MgO (001) substrate. Because the gold does not oxidize significantly, only a single period is observed. (Cole et al. 2004). (Reproduced with permission of the Institute of Physics.)

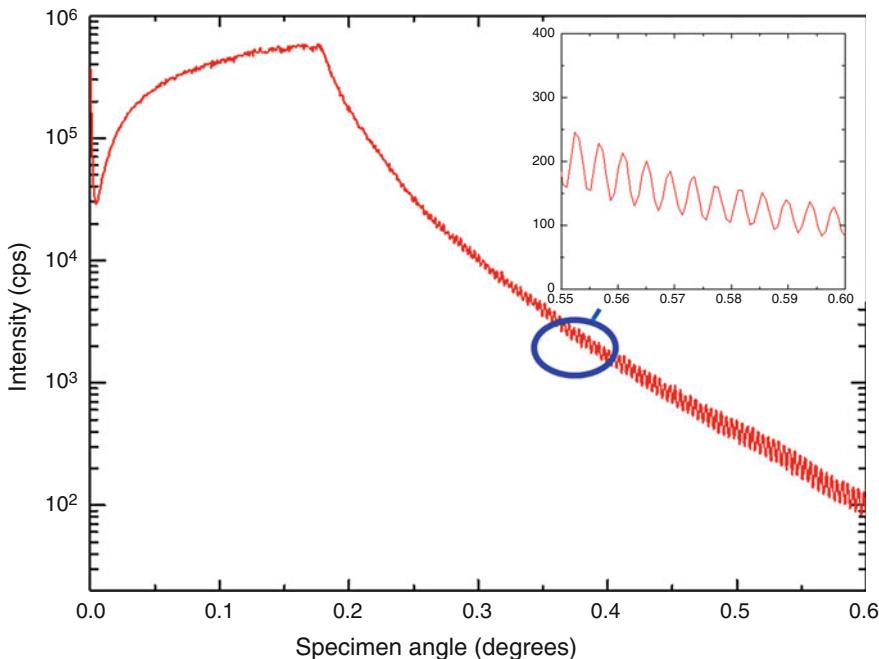


**Fig. 15** Single period interference pattern from a single 20 nm film of amorphous InGaZnO on a silicon substrate. (From Troughton et al. 2017, reproduced by permission of the American Institute of Physics)

plotting data on a log-linear scale.) Spin-coated organic single polymer films also often show very simple single-period interference patterns.

The limit on the maximum thickness of the film that can be measured by X-ray specular reflectivity is set by the period and amplitude of the interference fringes. As the film becomes thicker, so the period decreases. Very high-resolution X-ray optics are required, and a very small step size, and hence many data points, is required to resolve the fringes. Coupled with the fact that the amplitude of the fringes decreases as the film thickness increases, a “rule of thumb” is that the technique is only applicable to films of less than 1  $\mu\text{m}$  thickness (Fig. 16).

Unlike gold, metal oxide or organic polymer examples given above, most semiconductor or metallic materials readily oxidize and X-ray reflectivity measurements show that an oxide layer of typically 2 nm thickness is formed at the surface almost instantaneously on exposure to air. In the case of GaAs, the oxide layer grows substantially as a function of time. It is best simulated as a separate layer with a relatively wide interface (Fig. 17). There, we note that the diffuse scatter, as measured by an off-specular  $\theta$ – $2\theta$  scan with the specimen displaced  $0.1^\circ$  from the specular condition, is generally at least two orders of magnitude less than the

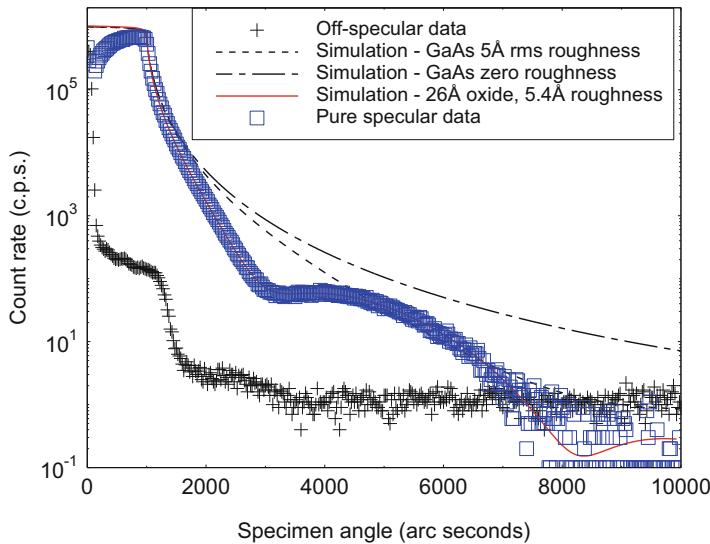


**Fig. 16** Specular reflectivity from a  $1.02\text{ }\mu\text{m}$  polymethyl methacrylate (PMMA) film showing very closely spaced, low-amplitude interference fringes

specular scatter until the incidence angle is about  $2^\circ$ , i.e., the scattering angle is about  $4^\circ$ . Therefore, the topological roughness is quite small.

We also note, in Fig. 17, that the measured reflectivity does not resemble the simple fall associated with a compositionally graded surface illustrated in Fig. 6. There is a well-defined single fringe. By fitting to a model of an oxide layer of different electron density with interfaces that change composition according to error functions, we can deduce that the oxide layer is  $2.6\text{ nm}$  thick. The precision on the measurement is set at  $0.1\text{ nm}$ , as it is possible to get equally good fits with changes to layer thickness and surface interface width of typically this value. To obtain a satisfactory fit, the width of Gaussian density gradient function that is the differential of the error function must be included at  $0.54\text{ nm}$ .

In the study of Tanner et al. (2001), the oxide thickness was observed to increase from about  $2\text{ nm}$  to  $3\text{ nm}$  over a period of almost 2 years after polishing. Additionally, there was an increase in the width over which the composition is graded. The study provided a clue as to why there is a time limit on the storage of epi-ready GaAs wafers, reinforced by optical photoabsorption analysis that showed that there were also changes in the oxide that made thermal desorption more difficult.

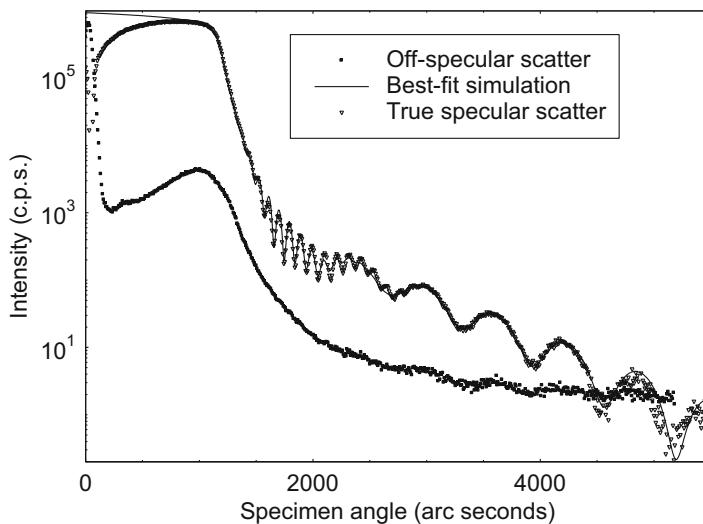


**Fig. 17** Grazing incidence reflectivity profile of a GaAs wafer immediately following polishing. The discrepancy between experiment and simulation in the region below the critical angle arises from the finite length and non-zero curvature of the sample. (Tanner et al. 2001). (Reproduced by permission of Elsevier.)

## Metrology of Multiple Films Using Specular Reflectivity

Where there are two films of different composition and thickness on the substrate, we observe two periods in the interference pattern, corresponding to the inverse thicknesses of each layer. As the oxide of most materials is significantly different in density compared with the native material (silicon being a notable exception), it is common to observe modulation of the interference pattern due to the thin oxide layer, typically 1–3 nm thick, at the surface of all materials exposed to air. The periods associated with each layer are not always visible in all parts of the interference profile. An illustration is given in Fig. 18 from work on indium tin oxide sputtered on glass that had a thin buffer layer of silicon dioxide deposited prior to the ITO deposition (Clarke et al. 2003). Because the ITO layer is relatively thick ( $112.03 \pm 0.02$  nm), at low angles, the X-ray wave does not penetrate to the silicon dioxide layer, the fringe pattern being of high frequency. As the incidence angle increases above about 3000 arc seconds, the wave penetrates the buffer layer, and as the high-frequency oscillations become weaker, the long-period oscillations associated with the  $22.32 \pm 0.02$  nm silicon dioxide layer dominate.

When more layers are included, we find a periodicity in the interference fringe pattern corresponding to each layer. For multiple layers, this can result in very



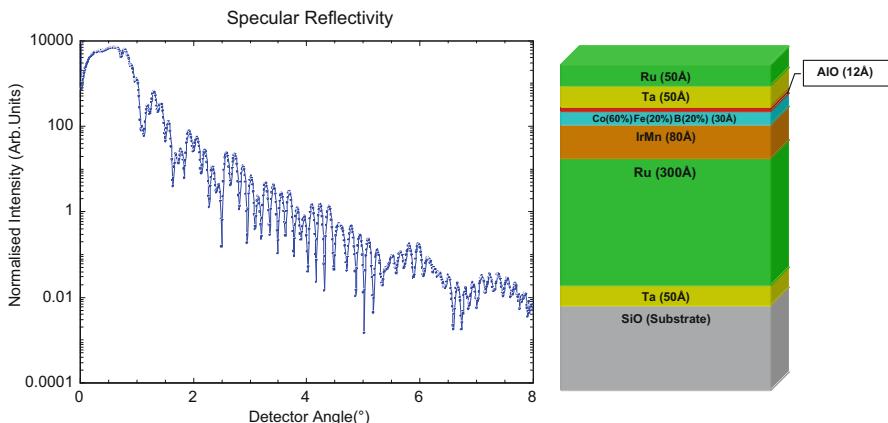
**Fig. 18** True specular and off-specular X-ray scatter from an indium tin oxide layer on a glass substrate with a thin silicon dioxide buffer layer. The off-specular data were collected as a  $\theta-2\theta$  scan with the specimen off-set  $-0.1^\circ$  from the specular condition. The true specular scatter is the measured specular scatter from which is subtracted the off-specular scatter. (Clarke et al. 2003). (Reproduced by permission of the Institute of Physics.)

complex patterns, as exemplified in Fig. 19. These data are from a common magnetic tunnel junction system which normally consists of two amorphous FeCoB electrodes separated by an insulating alumina barrier, which is sufficiently thin for spin-polarized electrons to tunnel across the junction. There is, in addition, a layer of antiferromagnetic IrMn alloy, used to pin the magnetization of the lower CoFeB layer and buffer layers of Ru and Ta in order to promote columnar growth. Figure 19 is of a structure with the top CoFeB electrode missing.

Despite the complexity, it was possible to show from changes in a very specific region of the reflectivity profile, that sharpening of the interface between the alumina and CoFeB layers occurs on annealing (Pym et al. 2006).

### Metrological Analysis of Reflectivity Profiles: How to Extract Material Parameters from the Data

As most nanotechnology manufacturing processes involve the deposition of more than one film, it is here that the true challenges lie. As we have seen, each film gives rise to a specific periodicity in the interference pattern and identification of these by some method is necessary to solve the problem. Although Fourier analysis has been attempted, it has severe limitations due partly to issues of signal-to-noise, dominance of different periods in different parts of the pattern (Fig. 18) and also that the fringe period is not exactly constant as the scattering weakens and the conditions under



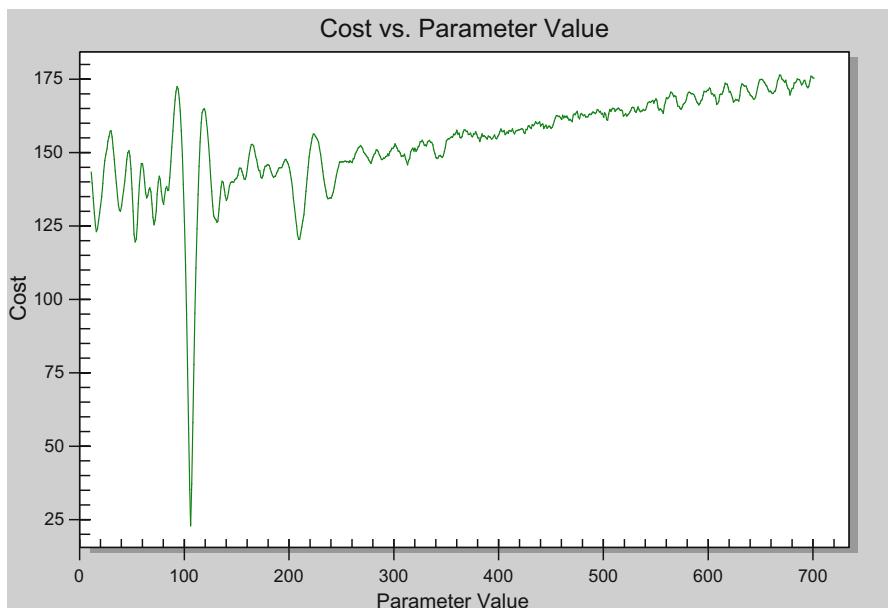
**Fig. 19** (a) True specular reflectivity profile from the tunnel junction structure shown in (b). Here the alumina barrier lacks its top CoFeB electrode. X-ray wavelength 1.3 Å. (From Pym et al. 2006, reproduced by permission of American Institute of Physics.)

which the Distorted Wave Born Approximation must be used gives way to the Born Approximation regime. Further, the Fourier approach does not enable the user to determine the density of individual films nor does it enable interface width to be determined.

The analytical approach adopted by all researchers and commercial vendors alike is that of matching data to the predictions of a structural model. Theoretical specular reflectivity profiles, using Parratt's (1954) approach, are generated from an approximate structure, which is then refined to minimize the difference between simulation and experiment. Commercial vendors usually include reflectivity fitting software as part of the instrument package, but Open Source software is available on the Web. Probably the most widely used package is GenX (Bjorck and Andersson 2007). The main problem with fitting complex experimental profiles to simulated ones is that of finding the global minimum in the intensity difference as opposed to local minima that do not represent the optimum fit of the model to the data. In the early days of data fitting, the downhill simplex algorithm was regularly used (e.g., Lucas et al. 1988), but this was found to work only if the thickness, density, and roughness were known quite accurately prior to X-ray metrology. Although there have been some reports of the use of simulated annealing to find the global minimum, GenX and most commercial vendors have followed the original approach of Wormington et al. (1999) and use genetic algorithms to identify the true minimum. Because it is the high-angle, weak scattering that provides the most precise information on layer thickness, it is important to minimize the difference between the logarithms of the intensities in the experimental and simulated data, not the absolute intensities. This strategy avoids overweighting the intense low-angle region, where systematic errors are greatest.

Figure 20 shows the “cost,” that is, the difference between simulated and experimental data on a log scale versus the variation of a particular layer thickness parameter. It is easy to identify the minimum where there is a best fit of the simulation to the experimental data. However, we should note that there are a number of local minima, well above the statistical noise level into which a simple “downhill” algorithm will get trapped. Even more challenging are the two large maxima that exist on either side of the deep minimum, so even starting the optimization with the parameter quite close to the best-fit position would not necessarily locate its best value.

The means by which the genetic algorithm and simulated annealing programs avoid the above trap is to explore parameter space more widely and jump to different regions in order to get out of the local minima. In addition to the seminal paper by Wormington et al. (1999), other important contributions to the use of genetic algorithms for fitting X-ray reflectivity data are found in the papers of Dane et al. (1998), Ulyanenkov and Sobolewski (2005), and Tiilikainen et al. (2007a, b). Simulated annealing was earlier used by Kunz et al. (1993) for neutron reflectivity fitting and has been used for X-ray reflectivity by Solookinejad et al. (2012). A comparison of the simulated annealing technique with genetic algorithms has been given by van der Lee et al. (2007), and they have concluded that the latter method is superior.



**Fig. 20** Example of the “cost,” i.e., the log difference of the total experimental and simulated intensities, as a function of the parameter value. The global minimum in the log difference is visible as a sharp minimum, with two large maxima on either side. This is characteristic of many fitting processes in which the layer thickness is being determined

In the case of magnetic thin film materials, use of both X-ray and polarized neutron specular reflectivity can be very powerful (Vaz et al. 2007), with the X-ray technique having a high resolution for the structural metrology, while the polarized neutron reflectivity yields layer by layer information on the magnetic moments. Co-refinement of neutron and X-ray reflectivity data has been demonstrated by Nelson (2006). Van der Lee et al. (2007) give a review of the complementarity of the X-ray and neutron techniques.

It is not possible to give a specific value for the precision to which film thickness can be determined. The precision is dependent on the film thickness. At nanometer film thickness, few fringes will be visible, and it is relatively difficult to identify the peak(s), as, for example, in Fig. 17. In that example, the precision was 100 pm in 2.6 nm. On the other hand, for the somewhat thicker ITO film of Fig. 18, the precision is 20 pm in 112 nm, because more fringes are visible and the data are better constrained. Indeed, the simulated and experimental data for the true specular scatter are extremely difficult to tell apart in Fig. 18. In the “sweet spot” of about 50–200 nm, a precision of 20 pm is a good rule of thumb for both film thickness and interface width. A further good example of the precision of this thickness region is from a 97.8 nm poly-methyl methacrylate (PMMA) film spun-coated on a silicon wafer (Fig. 21). Again, it is extremely difficult to distinguish experimental and simulated curves.

The absolute value of the precision does not improve as the film thickness increases due to the reduction in both the amplitude of the fringes and the fringe period becoming comparable with the instrument resolution and the step size for reasonable data collection times.

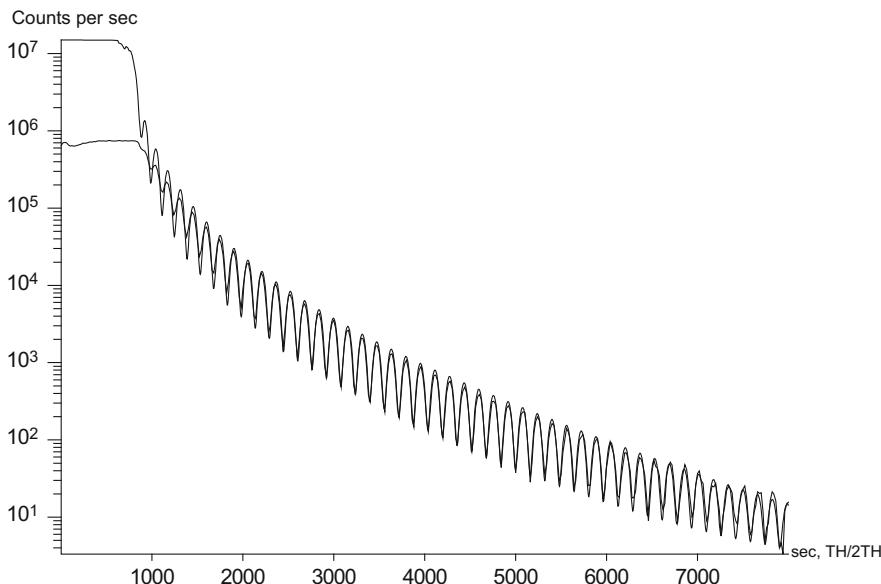
---

## NDE Using X-Ray Reflectivity in the Laboratory

Most analytical laboratories that use X-ray reflectivity as a thin film metrology do so with multipurpose high-resolution X-ray diffraction systems. The approach has not changed much from that originally employed by Parratt (1954) shown in Fig. 1. There is a need for:

1. Modest monochromatization of the X-ray beam using a crystal and slit arrangement
2. Collimation of the beam to ensure that the axis of rotation coincides with the position that the X-ray beam strikes the specimen
3. A low noise detector that can be scanned in a ratio of 2:1 with the specimen

A bent graphite monochromator has a wide mosaic spread, which in front of a narrow slit will provide a both modest wavelength dispersion and parallelism. The advantage is in intensity and dynamic range. Shindler and Suter (1992) showed that using this configuration, X-ray reflectivity data covering eight decades of intensity can be obtained with a rotating anode X-ray source. From a point source, a flat crystal monochromator results in a spectrum spread out in space. Use of a slit to



**Fig. 21** Reflected intensity versus incidence angle for a single poly-methyl methacrylate film spun on a silicon substrate. The best fit gave  $97.8 \text{ nm} \pm 20 \text{ pm}$  film thickness with  $0.45 \text{ nm}$  interface width at top and bottom film surfaces. Inclusion of a  $100 \text{ m}$  radius of curvature of the sample and  $25 \text{ s}$  instrument function was needed to match the data

select the  $\text{K}\alpha$  characteristic lines results in a parallel, well-monochromated beam. The curved crystal monochromator results in a nearly parallel, quasi-monochromatic beam across the whole output width of the beam. High intensity, typically up to  $10^8$  X-ray photons per second in the reflected beam can be obtained by use of a parabolically curved, graded multilayer mirror. These mirrors are known as Goebel mirrors (Schuster and Goebel 1995, 1996; Michaelsen et al. 2000; Schuster et al. 2001) and deliver a beam parallel to better than  $0.05^\circ$  with dispersion sufficiently large to separate the  $\text{CuK}\alpha$  and  $\text{CuK}\beta$  lines. Most systems fitted with Goebel mirrors also use a flat Ge monochromator to reduce the wavelength band passed to the sample.

The beam width used depends on the length of the sample. As X-ray reflectivity is performed at grazing incidence, the beam footprint on the sample is large and therefore the technique is not able to select small areas for examination in the same way as can be done in high-resolution X-ray diffraction. In any case, a narrow slit in front of the crystal is necessary in order to restrict the footprint size. An alternative strategy is to use a knife edge above the sample surface in order to restrict the area from which scattering occurs. The large footprint also implies that the sample must be flat and not figured. It must normally be larger than typically 5 mm in length. (Expert users can handle samples down to 2 mm in length but, it is challenging!)

A key requirement is that the sample rotation axis coincides with the point that the beam strikes the surface of the sample. The alignment procedure for a  $\theta-2\theta$  diffractometer, where the sample and detector move, has been described by the present author in another Springer handbook (Tanner 2015). An alternative diffractometer geometry is the so-called  $\theta-\theta$  geometry, where the sample is set horizontal and does not move, but both source and detector move at equal rates. The advantage of this geometry is that no adhesive or vacuum arrangement is needed to hold the sample, as is the case in the  $\theta-2\theta$  geometry. Such a setting works well also in the context of robotic handling, as well as when liquid is above the surface.

Suitable high-resolution diffractometers on which X-ray reflectivity data of high quality can be taken are available from a range of commercial manufacturers, including Bruker, PANalytical, and Rigaku.

## NDE in the Fab

The full-scale NDE application of X-ray reflectivity in a high volume manufacturing environment is still in its infancy. As with all in-line and in-fab metrology, the key challenge is speed of analysis. With a laboratory-based high-resolution diffractometer incorporating a standard X-ray tube, a reflectivity profile can be recorded in 10–20 min, depending on the range to be scanned and the information required of the film thickness and interface width. If data are required out to an incidence angle of a few degrees, it may take several hours to complete the measurement.

The key to all high-speed NDE methods is to collect data in parallel, rather than serially, and X-ray reflectivity is no exception. Here, the challenge is to design a monochromator that will deliver a range of incidence angles simultaneously and to have a detector of sufficient aperture and dynamic range that the full reflectivity profile can be recorded simultaneously. The measurement time will then be limited by the Poisson statistics of the highest angles from which information is needed.

One route to parallel data collection is to use a parallel beam of white radiation, with specific incidence and exit angles, and record the scattering as a function of energy using an energy dispersive detector. As the modulus of the wavevector  $\mathbf{k}_I$  is related to the energy  $E$  by

$$|\mathbf{k}_I| = E/hc, \quad (19)$$

for the fixed angle, the scattering vector  $\mathbf{Q}$  varies systematically with energy. As the reflectivity profile is fundamentally determined by the scattering vector, and not by angle, a reflectivity curve is obtained. This approach was tried in the 1980s with laboratory X-ray sources but due to the intense characteristic lines saturating the detector, it did not prove totally satisfactory (Tanner et al. 1991; Chason and Warwick 1991). Subsequent in situ laboratory-based studies, where the range of wavevector was limited, have been more successful (Chason et al. 1992; Paci et al. 2006). The data collection time is typically a few to tens of minutes using the continuous component of the spectrum (Bremstrahlung) from a laboratory X-ray

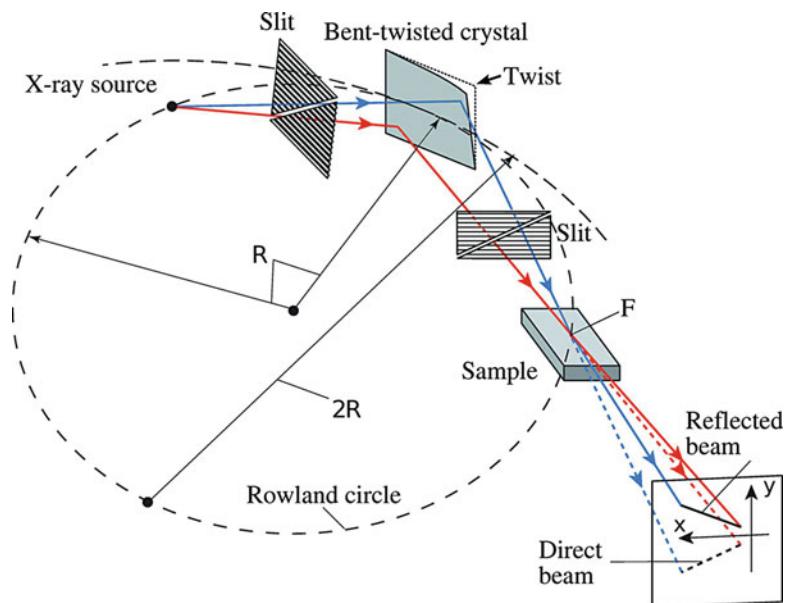
source. Higher speed can be obtained by using the continuous spectrum and parallel beam from a synchrotron radiation source. For example, using energy dispersive detection at the BESSY synchrotron radiation source, reflectivity profiles of polymer films were recorded in less than a second (Bhattacharya et al. 2003). Although there have been a number of experiments reported using this technique, particularly for in-situ studies, it has not become widely adopted by the research or NDE communities.

The most successful approach to high-speed reflectivity has been to use a diverging or converging monochromatic beam, (the angular divergence covering the full range of angles to be measured) using, a linear or two-dimensional detector. As with energy-dispersive detection, there is no need to rotate the sample, nor the detector, nor the X-ray source. Three decades ago, Naudon et al. (1989) described an apparatus to achieve this using characteristic X-rays from a line focus laboratory X-ray source, a knife edge close to the sample surface, and a one-dimensional detector. The technique, despite its aberrations (Stoev and Sakurai 2011), was exploited by Mizusawa and Sakurai (2011) to study in situ the changes in thickness of a 100 nm polyvinyl acetate film as it passed through the glass transition temperature. Measurement times were typically 0.1–20 s.

Using white synchrotron radiation, Voegeli et al. (2013) and Arakawa et al. (2013) succeeded in recording full reflectivity profiles on the millisecond timescale using curved crystal monochromators. In the latter paper, the group used a bent and twisted crystal to focus the X-rays in such a way that as a function of direction, both the incidence angle and wavelength change at the specimen. Recently, the group has described how this design can be adapted to work with characteristic X-rays from a laboratory point source (Voegeli et al. 2017). Here, the bent and twisted crystal monochromatizes and focuses the diverging characteristic CuK $\alpha$  X-rays from a laboratory point source onto the sample (Fig. 22). With a two-dimensional detector, the specularly reflected X-rays can be distinguished from the diffuse scattering, enabling a large dynamic range to be obtained. With an Au film on Si a true reflectivity profile down to the order of  $10^{-6}$  was obtained in 100 s and  $10^{-5}$  in a measurement time of 10 s (Voegeli et al. 2017).

Two companies market fully automated in-fab tools for X-ray reflectivity NDE applications on wafers up to 300 mm, namely, Bruker and Rigaku (Fig. 23). Neither company is willing to disclose its monochromator technology or technique. However, as Bruker purchased Jordan Valley Semiconductors, which now forms its Semiconductor Division, we may infer the general principles of operation of a high-speed X-ray reflectivity tool from the Jordan Valley Semiconductors patents (Yokhin et al. 2003; Berman and Dikopoltsev 2004).

The geometry described in the main patent (Fig. 24) appears to be similar to that of Naudon et al. (1989), which is cited in the patent. In Fig. 24, item 26 comprises the monochromator assembly, which delivers a converging beam of X-rays focused at the sample. There are thus a range of incidence angles reflected from that point and by use of a position sensitive detector, the reflected intensity at each point is directly related to the incidence angle. A full reflectivity profile is thus recorded. In the patent, the preferred optics “comprise a curved crystal monochromator, such as

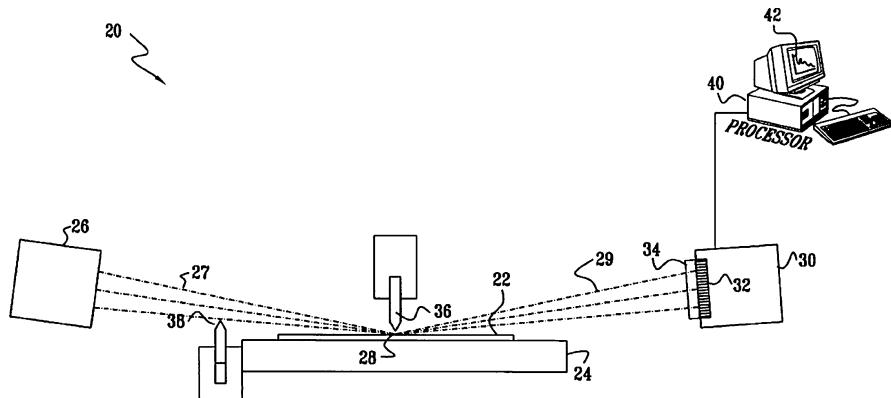


**Fig. 22** High-speed reflectometer of Voegeli et al. (2017). (Reproduced with permission of the International Union of Crystallography.)

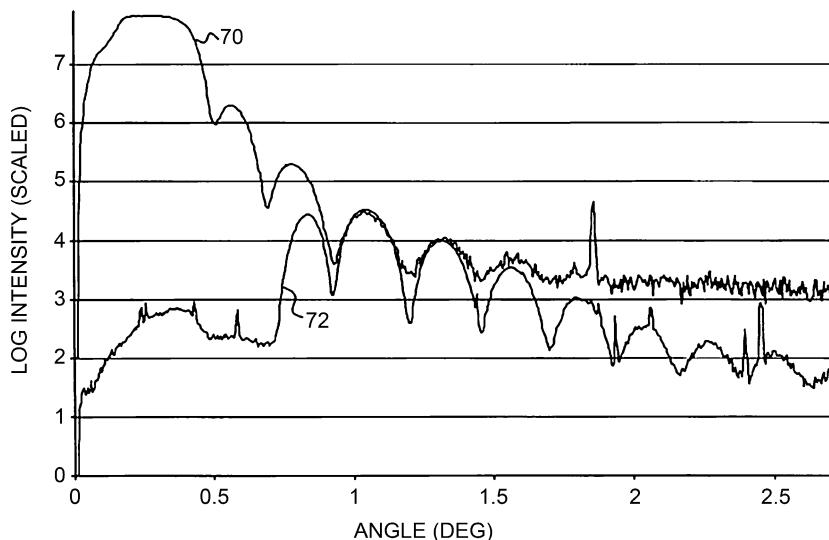


**Fig. 23** (a) The Rigaku MFM310 and (b) Bruker JVX7300 robotically loaded X-ray reflectivity tools. (Reproduced with permission.)

the Doubly-Bent Focusing Crystal Optic, produced by XOS Inc., of Albany, NY.” They do, nevertheless, note that other curved crystal monochromators may be used. However, because of the focusing in the two orthogonal directions, the doubly bent monochromator system does minimize the area of the sample illuminated, potentially allowing patterned wafers to be measured.



**Fig. 24** The geometry of the reflectometer described in the Jordan Valley patent. (Yokhin et al. 2003)



**Fig. 25** Reflectivity profiles recorded with the dynamic shutter allowing the full beam to impinge on the sample (upper curve) and with the shutter cutting out the low angle of incidence X-rays (lower curve). (Yokhin et al. 2003)

A novel feature of the system described is a dynamic shutter, which in its low-angle position, allows the entire beam of X-rays from the X-ray source to impinge on the sample, while in the high-angle position, the shutter blocks low-angle X-rays thereby extending the dynamic range (Fig. 25), as the background scatter is substantially reduced. By stitching the two curves together, a full profile is obtained.

## Conclusions

As is evident from the above discussion, the principles of X-ray reflectivity are well understood and analysis of data is well developed. The method has been used as a laboratory technique for nondestructive evaluations for many years, and standard commercial high-resolution X-ray diffractometers are all well adapted for such measurements. The precision of thickness measurement is not constant but varies with film thickness. For films in the “sweet spot” of typically 100 nm, a precision of 20 pm is possible (Clarke et al. 2003). Despite its extensive laboratory and synchrotron radiation source use, the technique is in many ways still in its infancy as a metrology process for industrial nondestructive evaluation. However, it is apparent that within a niche in semiconductor manufacturing, the method is extremely powerful and its use can only be set to increase.

---

## References

- Allain M, Benattar JJ, Rieutord F, Robin P (1987) Surface study of Langmuir-Blodgett films by electron microscopy and X-ray reflectivity. *Europhys Lett* 3:309–314
- Anastasiadis SH, Russell TP, Satija SK, Majkrzak CF (1989) Neutron reflectivity studies of the surface-induced ordering of diblock copolymer films. *Phys Rev Lett* 62:1852–1855
- Arakawa E, Voegeli W, Matsushita T, Yano YF, Hatano T (2013) Quick X-ray reflectometry in the simultaneous multiple angle-wavelength dispersive mode. *J Phys Conf Ser* 425:092002
- Baribeau JM (1993) X-ray reflectometry study of interdiffusion in Si/Ge heterostructures. *J Appl Phys* 74:3805–3810
- Berman D, Dikopoltsev A (2004) X-ray reflectometry of thin film layers with increased accuracy. US Patent Application No. US 2004/0131151 A1
- Bhattacharya M, Mukherjee M, Sanyal M, Geue T, Grenzer J, Pietsch U (2003) Energy dispersive x-ray reflectivity technique to study thermal properties of polymer films. *J Appl Phys* 94:2882–2887
- Bjorck M, Andersson G (2007) GenX: an extensible X-ray reflectivity refinement program utilizing differential evolution. *J Appl Cryst* 40:1174–1178
- Blundell SJ, Bland JAC (1992) Polarized neutron reflection as a probe of magnetic films and multilayers. *Phys Rev B* 46:3391–3400
- Bosio L, Benattar JJ, Rieutord F (1987) X-ray reflectivity of a Langmuir monolayer on water. *Rev Phys Appl* 22:775–778
- Bowen DK, Tanner BK (2006) X-ray metrology in semiconductor manufacturing. CRC Press, Taylor and Francis, Boca Raton, p 279
- Chason E, Mayer TM (1997) Thin film and surface characterization by specular X-ray reflectivity. *Crit Rev Solid State Mater Sci* 22:1–67
- Chason E, Warwick DT (1991) X-ray reflectivity measurements of surface roughness using energy dispersive detection. *Mater Res Soc Symp Proc* 208:151–356
- Chason E, Mayer TM, Payne A, Wu D (1992) In-situ energy dispersive X-ray reflectivity measurements of H ion bombardments on SiO<sub>2</sub>/Si and Si. *Appl Phys Lett* 60:2353–2355
- Chen H, Heald MJ (1989) Concentration profiling using X-ray reflectivity – application to Cu-Al interfaces. *J Appl Phys* 66:1793–1799
- Clarke J, Pape I, Normile P, Tanner BK (2003) X-ray scattering from uniform and patterned indium tin oxide thin films. *J Phys D Appl Phys* 36:A209–A213

- Cole A, Hickey BJ, Hase TPA, Buchanan JDR, Tanner BK (2004) Influence of the interfacial roughness on the electron channelling in Fe/Au(001) multilayers. *J Phys Condens Matter* 16:1197–1209
- Cubitt R, Fragneto G (2002) D17: the new reflectometer at the ILL. *Appl Phys A Mater Sci Process* 74:S329–S331
- Dane AD, Veldhuis A, de Boer DKG, Leenaers AJG, Buydens LMC (1998) Application of genetic algorithms for characterization of thin layered materials by glancing incidence X-ray reflectometry. *Physica B* 253:254–268
- De Boer DKG (1994) Influence of the roughness profile on the specular reflectivity of X-rays and neutrons. *Phys Rev B* 49:5817–5820
- De Boer DKG (1996) X-ray scattering and X-ray fluorescence from materials with rough interfaces. *Phys Rev B* 53:6048–6064
- Ferrari AC, Li Bassi A, Tanner BK, Stolojan V, Yuan J, Brown LM, Rodil SE, Kleinsorge B, Robertson J (2000) Density,  $sp^3$  fraction, and cross-sectional structure of amorphous carbon films determined by X-ray reflectivity and electron energy-loss spectroscopy. *Phys Rev B* 61:11089–11103
- Huang TC, Nozieres JP, Speriosu VS, Lefakis H, Gurney BA (1992) X-ray reflectivity analysis of giant-magnetoresistance spin-valve layered structures. *Appl Phys Lett* 60:1573–1575
- Huang TC, Nozieres JP, Speriosu VS, Gurney BA, Lefakis H (1993) Effect of annealing on the interfaces of giant-magnetoresistance spin-valve structures. *Appl Phys Lett* 62:1478–1480
- Hudson JM, Tanner BK (1992) Characterization of magnetic multilayers by grazing incidence X-ray reflectivity. *IEEE Trans Magn* 28:2736–2741
- Hudson JM, Powell AR, Bowen DK, Wormington M, Tanner BK, Kubik RA, Parker EHC (1992) Thermal degradation of SiGe interfaces studied by X-ray reflectivity and diffraction. *Mater Res Soc Symp Proc* 239:455–460
- Isherwood BJ (1977) Characterization of thin-films by X-ray diffraction. *GEC J Sci Technol* 43:111–124
- Kiessig H (1930) Interferenz von Röntgenstrahlen an dünnen Schichten [Interference of X-rays on thin layers]. *Naturwissenschaften* 18:847–848
- Kiessig H (1931a) Untersuchungen zur total reflexion von Röntgenstrahlen. *Ann Phys* 10:715–768
- Kiessig H (1931b) Interferenz von Röntgenstrahlen an dünnen Schichten. *Ann Phys* 10:769–788
- Koenig BW, Kruger S, Orts WJ, Majkrzak CF, Berk NF, Silverston JV, Gawrisch K (1996) Neutron reflectivity and atomic force microscopy studies of a lipid bilayer in water adsorbed to the surface of a silicon crystal. *Langmuir* 12:1343–1350
- Kunz K, Reiter J, Gotzelmann A, Stamm M (1993) Model-free analysis of neutron reflectivity data from polymer thin-films with the simulated annealing technique. *Macromolecules* 26: 4316–4323
- Lewicky R, Herne TM, Tarlov MJ, Satija SK (1998) Using self-assembly to control the structure of DNA monolayers on gold: a neutron reflectivity study. *J Am Chem Soc* 120:9787–9792
- LiBassi A, Ferrari AC, Stolojan V, Tanner BK, Robertson J, Brown LM (2000) Density,  $sp^3$  content and internal layering of DLC films by X-ray reflectivity and electron energy loss spectroscopy. *Diamond Relat Mater* 9:771–776
- Lucas CA, Hatton PD, Bates S, Ryan TW, Miles S, Tanner BK (1988) Characterization of nanometer-scale epitaxial structures by grazing-incidence X-ray diffraction and specular reflectivity. *J Appl Phys* 63:1936–1941
- Martinbouyer G, Dethy B (1982) Determination of the thickness of thin-films by the Kiessig fringe technique. *Analisis* 10:234–238
- Meekins JF, Cruddance RG, Gursky H (1986) Optimization of layered synthetic microstructures for narrowband reflectivity at soft X-ray and EUV wavelengths. *Appl Optics* 25:2757–2763
- Michaelsen C, Ricardo P, Anders D, Schuster M, Schilling J, Goebel H (2000) Improved graded multilayer mirrors for XRD applications. *Adv X-Ray Anal* 42:308–320
- Mizusawa M, Sakurai K (2011) In-situ X-ray reflectivity measurement of polvinyl acetate thin films during glass transition. *IOP Conf Ser Mater Sci Eng* 24:012013

- Naudon A, Chihab J, Goudeau P, Mimault J (1989) New apparatus for grazing X-ray reflectometry in the angle-resolved dispersive mode. *J Appl Cryst* 22:460–464
- Nelson A (2006) Co-refinement of multiple-contrast neutron/X-ray reflectivity data using MOTOFIT. *J Appl Cryst* 39:273–276
- Paci B, Generosi A, Rossi Albertini V, Perfetti P, de Bettignies R, Leroy J, Firon M, Sentein C (2006) Controlling photoinduced degradation in plastic photovoltaic cells: a time-resolved energy dispersive X-ray reflectometry study. *Appl Phys Lett* 89:043507
- Pape I, Tanner BK, Wormington M (1999) Grazing incidence X-ray scattering measurement of silicate glass surfaces. *J Non-Cryst Solids* 248:75–83
- Pape I, Lawrence CW, Roberts SG, Briggs GAD, Kolosov OV, Hey AW, Paine CF, Tanner BK (2000) Evaluation of polishing damage in alumina. *Philos Mag A* 80:1913–1934
- Parratt LG (1954) Surface studies of solids by total reflection of X-rays. *Phys Rev* 95:359–369
- Penfold J (1991) Instrumentation for neutron reflectivity. *Physica B* 173:1–10
- Penfold J (2002) Neutron reflectivity and soft condensed matter. *Curr Opin Colloid Interface Sci* 7:139–147
- Pym ATG, Lamperti A, Tanner BK, Dimopoulos T, Rührig M, Wecker J (2006) Interface sharpening in CoFeB magnetic tunnel junctions. *Appl Phys Lett* 88:162505
- Regan MJ, Kawamoto EH, Lee S, Pershan PS, Maskil N, Deutsch M, Magnussen OM, Ocko BM, Berman LE (1995) Surface layering in liquid gallium. *Phys Rev Lett* 75:2498–2501
- Rosen DL, Brown D, Gilfrith J, Burkhalter P (1988) Multilayer roughness evaluated by X-ray reflectivity. *J Appl Cryst* 21:136–144
- Satchell N, Witt JDS, Burnell G, Curran PJ, Kinane CJ, Charlton TR, Langridge S, Cooper JFK (2017) Probing the spiral magnetic phase in 6 nm textured erbium using polarised neutron reflectometry. *J Phys C Condens Matter* 29:055801
- Schuster M, Goebel H (1995) Parallel-beam coupling into channel-cut monochromators using curved graded multilayers. *J Phys D Appl Phys* 28:A270–A275
- Schuster M, Goebel H (1996) Calculation of improvement to HRXRD system through-put using curved graded multilayers. *J Phys D Appl Phys* 29:1677–1679
- Schuster M, Goebel H, Michaelsen C, Bormann R (2001) X-ray analysis apparatus with a graded multilayer mirror. US Patent US6226349
- Shindler JD, Suter RM (1992) Moderate resolution X-ray reflectivity. *Rev Sci Instrum* 63: 5343–5347
- Sinha SK, Sirota EB, Garoff S, Stanley HB (1988) X-ray and neutron scattering from rough surfaces. *Phys Rev B* 38:2297–2311
- Soles CL, Lee HJ, Hedden RC, Liu DW, Bauer BJ, Wu WL (2003) X-ray porosimetry as a metrology to characterize the pore structure of low-k dielectric films. In: Seiler DG, Diebold AC, Shaffner TJ, McDonald R, Zollner S, Khosla RP, Secula EM (eds) Characterization and metrology for ULSI technology. AIP conference series 683. American Institute of Physics, Melville, pp 576–580
- Solookinejad G, Rozatian ASH, Habibi MH (2012) Investigation of sol-gel grown ZnO thin film: wavelet analysis and simulated annealing optimized X-ray reflectivity. *In J Mod Phys* 26: 1250070
- Stevens KH, Braueninger H, Kaase H, Metzdorf J (1986) Reflectivity measurements in the vacuum ultraviolet wavelength range on technical surfaces for the Wolter I telescope on board the X-ray astronomy satellite ROSAT. *Astrophys Space Sci* 125:169–174
- Stoev K, Sakurai K (2011) Aberration effects in quick X-ray reflectivity of curved samples. *IOP Conf Ser Mater Sci Eng* 24:012014
- Tanner BK (2015) X-ray scattering from spintronic structures. In: Xu Y, Awschalom DD, Nitta J (eds) Handbook of spintronics. Springer Science and Business Media, Dordrecht, pp 919–945
- Tanner BK, Miles SJ, Bowen DK, Hart L, Loxley N (1991) X-ray reflectometry from semiconductor surfaces and interfaces. *Mater Res Soc Symp Proc* 208:345–350
- Tanner BK, Allwood DA, Mason NJ (2001) Kinetics of native oxide growth on epiready GaAs. *Mater Sci Eng B* 80:99–103

- Tiilikainen J, Bosund V, Mattila M, Hakkarainen T, Sormunen J, Lipsanen H (2007a) Fitness function and nonunique solutions in X-ray reflectivity curve fitting: crosserror between surface roughness and mass density. *J Phys D Appl Phys* 40:4259–4263
- Tiilikainen J, Bosund V, Tilli J-M, Sormunen J, Mattila M, Hakkarainen T, Lipsanen H (2007b) Genetic algorithm using independent component analysis in X-ray reflectivity curve fitting of periodic layer structures. *J Phys D Appl Phys* 40:6000–6004
- Toney MF, Brennan S (1989) Measurements of carbon thin-films by X-ray reflectivity. *J Appl Phys* 66:1861–1863
- Troughton JG, Downs P, Price R, Atkinson D (2017) Densification of a-IGZP with low-temperature annealing for flexible electronics applications. *Appl Phys Lett* 110:011903
- Ulyanenkov A, Sobolewski S (2005) Extended genetic algorithm: application to X-ray analysis. *J Phys D Appl Phys* 38:A235–AA23
- van der Lee AD, Salah F, Harzallah B (2007) A comparison of modern data analysis methods for X-ray and neutron specular reflectivity data. *J Appl Cryst* 40:820–833
- Vaz CAF, Lauhoff G, Bland JAC, Langridge S, Bucknall D, Penfold J, Clarke J, Halder SK, Tanner BK (2007) Interface dependent magnetic moments in Cu/Co,Ni/Cu/Si(001) epitaxial structures. *J Magn Magn Mater* 313:89–97
- Veldhuis SA, Brinks P, Stawski TM, Gobel OF, ten Elshof JE (2014) A facile method for the density determination of ceramic thin films using X-ray reflectivity. *J Sol-Gel Sci Technol* 71:118–128
- Voegeli W, Matsushita T, Arakawa E, Shirasawa T, Takahashi T, Yano YF (2013) A method for measuring the specular X-ray reflectivity with millisecond time resolution. *J Phys Conf Ser* 425:092003
- Voegeli W, Kamezawa C, Arakawa E, Yano YF, Shirasawa TA, Takahashia T, Matsushita T (2017) A quick convergent-beam laboratory X-ray reflectometer using a simultaneous multiple-angle dispersive geometry. *J Appl Cryst* 50:570–575
- Wainfan N, Parratt LG (1960) X-ray reflection studies of the anneal and oxidation of some thin solid films. *J Appl Phys* 31:1331–1337
- Wainfan N, Scott NJ, Parratt LG (1959) Density measurements of some thin copper films. *J Appl Phys* 30:1604–1609
- Wormington M, Pape I, Hase TPA, Tanner BK, Bowen DK (1996) Evidence of grading at polished surfaces from grazing incidence X-ray scattering. *Philos Mag Lett* 74:211–216
- Wormington M, Panaccione C, Matney KM, Bowen DK (1999) Characterization of structures from X-ray scattering data using genetic algorithms. *Philos Trans R Soc Lond A* 357:2827–2848
- Woronick SC, Yang BX, Krol A, Kao YH, Munekata H, Chang LL, Phillips JC (1987) X-ray reflectivity of InAs/GaAs heterostructures with surface and interface roughness. *J Phys* 48: 51–56
- Yokhin B, Dikopoltsev A, Mazor I, Berman D (2003) X-ray reflectometer. US Patent No. US 6,512,814 B2
- Yoneda Y (1963) Anomalous surface reflection of X-rays. *Phys Rev* 131:2010–2013
- You H, Chiarello RP, Kim HK, Vandervoort KG (1993) X-ray reflectivity and scanning tunnelling microscopy study of kinetic roughening of sputter deposited gold films during growth. *Phys Rev Lett* 70:2900–2903
- Zhang R, Somasundaran P (2006) Advances in adsorption of surfactants and their mixtures at solid/solution interfaces. *Adv Colloid Interface Sci* 123:213–229

---

## **Part VI**

### **Particle Methods**



# Neutron Radiography and Tomography

34

Wolfgang Treimer

## Contents

Introduction .....	1218
Neutron Index of Refraction .....	1219
Interactions of Neutrons with Matter .....	1222
Basics of Neutron Radiography and Tomography .....	1223
Geometries Used in Radiography and Tomography .....	1224
Resolution Functions (Gauss, LSF, MTF) .....	1226
Mathematics of Computerized Tomography .....	1230
Projection-Slice Theorem .....	1232
Theory of Back-Projection .....	1236
Fundamentals of Radiography and Tomography with Polarized Neutrons .....	1243
Some Historical Remarks .....	1243
Polarizing Neutrons .....	1245
Determination of Beam Polarization .....	1248
Interaction with Magnetic Fields .....	1249
Calculation of Magnetic Fields .....	1251
Neutron Polarization .....	1252
Reconstruction Methods .....	1254
Reconstruction of Continuous Magnetic Fields .....	1260
Polarized Neutron Imaging: Radon Transform for Image Reconstruction .....	1261
Experiments with Polarized Neutron Imaging .....	1266
Phase-Based Imaging Signals .....	1274
Some Historical Aspects and Basics .....	1274
Radiography and Tomography: Refraction .....	1281
Radiography and Tomography: Small Angle Scattering .....	1282
Experimental Results: Refraction and Small Angle Scattering .....	1284
Neutron Grating Interferometry .....	1289

---

W. Treimer (✉)

Department Mathematics-Physics-Chemistry, University of Applied Sciences Beuth Hochschule  
für Technik Berlin, Berlin, Germany  
e-mail: [treimer@beuth-hochschule.de](mailto:treimer@beuth-hochschule.de)

---

Bragg Edge Imaging Using Energy Selective Neutron Technique .....	1292
Theory .....	1292
Application .....	1294
References .....	1295

---

### Abstract

In this chapter, the basics and applications of neutron radiography and neutron tomography are presented, outlining the principles of neutron physics. Then in the subsequent sections, different methods of neutron imaging are described and examples are given.

---

## Introduction

Radiography and tomography with X-rays are well known; however, radiography and tomography with neutrons are much less known. The long tradition of neutron radiography started quite soon after the discovery of the neutron by J. Chadwick in 1932 (Chadwick 1932). Only 3 years later, in 1935, Hartmut Kallmann and Ernst Kuhn in Berlin (Germany) reported the first experiments on neutron radiographies. However, neutron scattering soon became increasingly important, when in the 1940s in the United States and Canada, high flux reactors were built and scattering instruments installed. Neutron radiography became a rather exotic method for several reasons, first because scattering methods were superior if location of atoms or crystal structures in matter had to be identified, then because inelastic scattering became famous measuring phonons or magnetic Bragg scattering, and finally because the length scale of interest was well below  $\mu\text{m}$ ; all of which could not be measured by means of neutron radiography.

The detection in neutron radiography is based on converting neutrons into gammas by means of so-called converter foils together with (special) X-ray films, whereby the spatial resolution of approximately  $50 \mu\text{m}$  was below the one of available CCD cameras for a very long time. Therefore, neutron radiography was reduced to irradiation imaging, which was of certain interest to quality control in several industrial applications.

It was nearly 50 years before the first high resolution neutron radiographies and tomographies (Schillingen 1996; Schillinger et al. 1996; Brenizer 2013), and nearly 10 years before polarized neutrons were used for neutron imaging (Treimer et al. 2005a, b, 2012a, b; Kardjilov et al. 2008).

Soon, several imaging techniques were developed that opened a wide range of different applications in materials science and fundamental physics.

Besides absorption tomography, refraction, small angle scattering, dark field, grating interferometer, Bragg edge, stress and strain, and radiography and tomography with polarized neutrons are the most used techniques to date.

Therefore, it was only a question of time until neutron radiography and tomography became an important tool for industrial use, i.e. nondestructive testing and for other

topics. In order to detect small defects in material, etc., the spatial resolution must fit the challenges of industrial fabrication processes. Very often pure absorption (attenuation) images do not meet the requirements of, for example, stress and strain studies of materials. Different methods of investigations had to be found, such as Bragg edge techniques (energy selective radiography and tomography) or phase contrast radiography, so that neutron radiography and tomography could become accepted by the industrial user committee. Sometimes, experiments far off from neutron imaging contributed to effects in neutron radiography and tomography and helped to explain them. For example, in high resolution imaging ‘shadow-effects’ were observed, and their explanation was found by considering ultra small angle scattering experiments and coherence considerations in dynamical diffraction of thermal neutrons (Treimer et al. 2005a, b; Treimer and Feye-Treimer 2011). The preceding experiments on neutron optics in 1995/1996 (Kraft 1996; Schaper 1996; Ernst 1997) created the basis explaining physics, which were eventually called ‘phase contrast’. Refraction could be detected, only, using a high angular resolution double crystal diffractometer, and it could be proven that edges or phase boundaries cause shadow effects in high resolution imaging. Later, study was extended to spin dependent refraction by wedge-shaped magnetic fields, and several experiments using Zeeman splitting and high resolution double crystal diffractometry were worked out (Strobl et al. 2007). However, even though all these studies were necessary and could explain the physics, this technique (high resolution double crystal diffractometry) had to be further developed for industrial use.

In the past few years, phase contrast radiography and tomography have also gained increasing importance concerning the technical and industrial applications of energy selective radiography and tomography. In the future, they will become one of the most used methods in nondestructive testing, characterizing material defects and material components. The measurements and data evaluation are quite tricky and sophisticated and thus require knowledge on both topics: materials science and neutron radiography and tomography. The same holds for the use of polarized neutrons for nondestructive testing. The fields of applications focus upon investigations of magnetic fields, especially in superconductors type I and type II. A look inside a sample in the superconducting state became possible with polarized neutron imaging; no other technique can visualize and quantify trapped magnetic fields in superconducting matter. Thus, especially, high temperature superconductors (HTSC) will be the most interesting candidates in nondestructive testing using polarized neutron imaging.

---

## Neutron Index of Refraction

Some properties of neutrons are helpful to know as well as the parameters used in the following chapters. Therefore, a collection of neutron data is given in Table 1. Neutrons have a mass  $m$ , no charge but a magnetic moment  $\mu_n$ , which makes them ‘magnetic sensitive’, i.e. they interact with magnetic fields with theirs spins. ‘Born by nuclear fission’ neutrons must be moderated down to energies of much less than 1 eV, i.e. energies  $\sim 0.1 - 0.005$  eV  $\triangleq 0.09$  nm – 0.4 nm, in order to use them for investigations of condensed matter. Such thermal neutrons behave like

**Table 1** Fundamental properties of the neutron (NIST Neutron scattering length and cross sections; NIST Center for Neutron Research)

Mass	$m = 1.6749286(10) \cdot 10^{-27} \text{ kg}$
Charge	0
Spin	$\frac{1}{2}$
Planck's constant	$h = 6.6260755 \cdot 10^{-34} \text{ J s}$
Planck's constant in eV s	$h = 4.135667662 \cdot 10^{-15} \text{ eV s}$
Velocity of light	$c = 299792458 \text{ ms}^{-1}$
Avogadro (Loschmidt) constant	$N_A = 6.022140857 \cdot 10^{23} \text{ Mol}^{-1}$
Boltzmann constant	$k_B = 1.38064852 \cdot 10^{-23} \text{ J K}^{-1}$
Magnetic moment	$\mu_\eta = -1.91304273 \cdot \mu_N$
Nuclear magneton	$\mu_N = 5.050783699 \cdot 10^{-27} \text{ JT}^{-1}$
Gyromagnetic ratio	$\gamma_n = 1.83247172 \cdot 10^8 \text{ s}^{-1} \text{T}^{-1}$
Mean life time	$T_{\frac{1}{2}} = 885, 8 \pm 0, 9 \text{ s}$
Elementary charge	$eV = 1.602176620810^{-19} \text{ C}$
Neutron gyromagnetic ratio	$\gamma_n = 1.83247172 \cdot 10^8 \text{ s}^{-1} \text{T}^{-1}$
Larmor frequency	$\omega_L = \gamma_n B = 1.83247172 \cdot 10^8 \text{ s}^{-1}$

classical particles because their velocity is much less than the speed of light. A free neutron also behaves like a wave, with a wavelength  $\lambda$  having a kinetic energy  $E$ , which can be calculated with the de-Broglie equation (Eqs. 1 and 2) using values from Table 1. Typically, neutron wavelengths are of the order 0.1 nm ( $10^{-10} \text{ m}$ ), and the kinetic energy  $E$  of a free neutron can be calculated classically.

$$\lambda = \frac{h}{mv}, \quad k = \frac{2\pi}{\lambda} \quad (1)$$

$$E = \frac{\hbar^2 k^2}{2m} = \frac{h^2}{2m\lambda^2} = \frac{mv^2}{2} \quad (2)$$

Wave phenomena, such as reflection, refraction and diffraction are well known from light optics, and their laws can be applied to thermal neutrons because their velocity  $v \ll c$  ( $c$  = speed of light). The interaction of neutrons with matter is the interaction with nuclear potentials  $V(\mathbf{r}, t)$  described by the Schrödinger equation (SE) (Messiah 1999; Rauch and Werner 2000).

$$H\Psi(\mathbf{r}, t) = \left( -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}, t) \right) \Psi(\mathbf{r}, t) = i\hbar \frac{\partial \Psi(\mathbf{r}, t)}{\partial t} \quad (3)$$

In the case of isotropic homogeneous matter  $V(\mathbf{r}, t)$  can be treated as an effective optical potential, given as

$$V = \frac{2\pi\hbar^2}{m} b_c N \quad (4)$$

where  $b_c$  is the mean coherent scattering length of the medium and a measure of the amount of the interaction (potential) and thus scattering power,  $N$  is the atomic density/unit volume ( $N = N_A \cdot \rho/A$ ) and  $A$  is the atomic weight.

The solution (Eq. 3) is found with Ansatz for  $\Psi(\mathbf{r}, t) = \psi(\mathbf{r}) \cdot e^{-i\omega t}$  for the time-independent SE

$$\nabla^2 \psi(r) + \mathbf{K}^2(\mathbf{r}) \psi(r) = E \psi(r), \quad (5)$$

$\mathbf{K}$  is the wave vector of the neutron wave in the matter. From (Eq. 5) one gets the solution

$$\mathbf{K}^2(\mathbf{r}) = \frac{2m}{\hbar^2} [E - V(\mathbf{r})]. \quad (6)$$

From (light) optics, the index of refraction  $n = n(r)$  is known as the ratio of the velocities  $v/v_0$  (velocity in matter/velocity in free space). With (Eq. 6)  $n(r)$  becomes

$$n(r) = \frac{K(r)}{k} = \left[ 1 - \frac{V(r)}{E} \right]^{\frac{1}{2}}, \quad (7)$$

$k$  is the wave vector of the neutron wave in free space. The scattering length  $b_c$  in (Eq. 4), is usually a complex number, actually  $b = b' + i b''$ , where the imaginary part represents absorption and incoherent scattering. The index of refraction  $n = 1 - \delta + i\beta$ , with

$$\delta + i\beta = \lambda^2 \frac{N}{2\pi} \sqrt{b_c^2 - \left( \frac{\Sigma_r}{2\lambda} \right)^2} + i \frac{\Sigma_r N \lambda}{4\pi}. \quad (8)$$

$\Sigma_r$  is the total reaction cross section including absorption and incoherent scattering cross sections (Rauch and Werner 2000). For non-magnetic material (interaction) the index of refraction  $n = n_{\text{nucl}}$  is finally

$$n_{\text{nucl}} = 1 - \lambda^2 \frac{Nb_c}{2\pi} \quad (\text{using } V(r) = V_{\text{nucl}} \ll E). \quad (9)$$

A similar derivation yields the magnetic index of refraction, setting the interaction potential  $V(r) = V_{\text{magn}}(\mathbf{r}) = \mu \cdot \mathbf{B}(\mathbf{r})$

$$n_{\text{magn}} = 1 \pm \frac{\mu_n B}{2E}, \quad (V(r) = V_{\text{magn}}(r) \ll E). \quad (10)$$

Including magnetic interaction, the index of refraction thus becomes

$$n = 1 - \lambda^2 \left( \frac{Nb_{\text{coh}}}{2\pi} \mp \frac{m\mu B}{h^2} \right) \quad (11)$$

The index of refraction plays a substantial role for all phase-based interactions of neutrons with matter, i.e. total reflection, diffraction, Bragg scattering, etc., and therefore also for phase contrast radiography and tomography, showing refraction and small angle contrast.

## Interactions of Neutrons with Matter

For NDT, three interactions of neutrons with matter are relevant:

- (a) Scattering: elastic, inelastic, magnetic, coherent, incoherent
- (b) Absorption, excitation: neutron interacts with a nucleus and creates e.g. a  $\gamma$ -quantum
- (c) Change of beam polarization (spin state of neutrons)

All interactions are commonly described by so-called cross sections, referring to the particular interaction the neutrons undergo, and strongly depend on their energy ( $1/v$  – law in the case of absorption for neutron energy less than 0.5 eV). The interactions of neutrons with matter can be distinguished into macroscopic and microscopic cross sections. Knowing the cross sections is important to calculate the amount of interaction and thus the expected size of the signal. A much recommended data reference is the NIST Center for Neutron Research link ([NIST Neutron](#) scattering length and cross sections; [NIST Center for Neutron Research](#)). The cross section is called microscopic if it refers to a single nucleus. The dimension of  $\sigma = [\text{cm}^2]$  is usually given in barn =  $10^{-24} \text{ cm}^2$ . The differential cross section  $\frac{d\sigma}{d\Omega}$  is defined as

$$\frac{d\sigma}{d\Omega} = \frac{\text{number of interacting particles}/(\text{unit time} \times \text{unit cone } d\Omega)}{\text{number of incident particles}/(\text{unit time} \times \text{unit area} \times \text{unit cone } d\Omega)}.$$

Note, the dimension of  $[\frac{d\sigma}{d\Omega}] = [\text{area}]$ . Often, the so-called macroscopic cross section  $\Sigma$  is used if the attenuation of a beam has to be calculated.  $\Sigma = N \times \sigma$ ,  $N$  is the number of nuclei per unit volume (e.g.  $\text{cm}^3$ ) and is calculated as  $N = \frac{\rho}{A} N_A$ , where  $\rho$  = density of material,  $A$  = atomic weight of material and  $N_A$  the Loschmidt or Avogadro number (see Table 1). The dimension of  $\Sigma = [\text{unit length}]^{-1}$ .

For example, iron has a density  $\rho = 7.86 \text{ g cm}^{-3}$ , the microscopic cross section for absorption of iron is 2.56 barn and the gram atomic weight is 55.847 g; numbers from ([NIST Center](#) for Neutron Research). The macroscopic cross section  $\Sigma$  becomes  $0.217 \text{ cm}^{-1}$ . A beam attenuation can then be determined with

$$I = I_0 \cdot e^{-\Sigma z}, \quad (12)$$

where  $z$  is the path length in the material; thus,  $I \propto 0.805$ . For a material consisting of  $k$  different atoms,  $\Sigma$  becomes the weighted sum of the cross sections of the constituents

$$\Sigma = \sum_k N_k \sigma_k, \quad (13)$$

where  $N_k$  is the density of the nuclei of kind  $k$  and  $\sigma_k$  the corresponding microscopic cross section of the nuclei. From this equation the attenuation law can be expressed as

$$I(x) = I_0 \cdot e^{-\int_{\text{path}} \Sigma(z) dz} \quad . \quad (14)$$

The inverse of  $\Sigma$ ,  $1/\Sigma = \lambda_{\text{free}}$ , is the free path of neutrons in the material. We will find this formula (Eq. 14) again in section “[Mathematics of Computerized Tomography](#).” Referring to the absorption law (Eq. 14), one defines a ‘half-thickness’  $z_{\frac{1}{2}}$  as

$$z_{\frac{1}{2}} = \frac{\ln(2)}{\int_{\text{path}} \Sigma(z) dz} \quad . \quad (15)$$

Sometimes, it becomes necessary to consider scattering processes that involve a macroscopic scattering cross section  $\Sigma_s(z)$ , which is given by

$$\Sigma_s(z) = \sum_k N_k(z) \cdot \sigma_{s,k} \quad . \quad (16)$$

In the case of neutron radiography and tomography, attenuation due to absorption is of major importance, i.e.,  $\Sigma$  (Eq. 13) must be replaced by the absorption cross section  $\Sigma_a(z)$

$$\Sigma_a(z) = \sum_k N_k(z) \cdot \sigma_{a,k} \quad . \quad (17)$$

In order to take both interactions into account  $\Sigma$  becomes the (macroscopic) total linear cross section and must be replaced by the sum of  $\Sigma_t(z) = \Sigma_a(z) + \Sigma_s(z)$ . Scattering processes usually contribute to a blurring of an image and decrease sharpness of mapped structures; however, they are also used as imaging signals. The interaction with magnetic fields due to the neutron spin is given in detail in section “[Fundamentals of Radiography and Tomography with Polarized Neutrons](#).”

## **Basics of Neutron Radiography and Tomography**

Radiography produces a two-dimensional attenuation image of a body or sample. The measured intensity in a point of observation is the (attenuated) intensity of a pencil neutron (or X-ray) that transmitted the sample. The contrast stems from different absorbing (attenuating) components in the sample, where the attenuation is summed up along the path through the sample, i.e. there is no local information (from the path) in the point of observation. In order to recover information (e.g. position dependent

absorption), tomographic methods must be used. Historically, in neutron imaging, neutrons were recorded with X-ray films together with special converter foils (Gadolinium or Indium) yielding quite a high resolution of app. 50–30  $\mu\text{m}$  or even less, depending on the thickness of the converter foil. Nowadays, they have been replaced by electronic cameras based on CCD or multi channel plates (MCP) (Tremsin et al. 2011). Electronic cameras have several advantages in comparison with films, i.e. data recording, data collecting, data image processing and data storage.

Because of the simple geometry, radiography (and even tomography) needs only a few parameters to work with, but it is important to settle some definitions concerning geometry, resolution, attenuation, etc. The step from a two-dimensional image to a three-dimensional one is rather small, using some mathematical tools that have to be applied. From projections (radiographies), 2D slices are reconstructed and stacked to get a 3D image of the sample.

For nondestructive testing (NDT), as well as for all other investigations of matter using radiation, the (neutron or X-ray) beam must be specified by flux  $\Phi$  = number of neutrons (photons)/unit time  $x$  unit area, by wavelength  $\lambda$  and wavelength band ( $\Delta\lambda/\lambda$ ), beam divergence  $\phi$  and, in the case of polarized neutron radiography and tomography, the polarization  $P$  (or spin state in the case of neutrons) of the beam must be known (see section “[Determination of Beam Polarization](#)”).

The neutron flux  $\Phi$  from a research reactor usually follows the shape of a Maxwell spectrum due to the thermal equilibrium of neutrons with the moderator. The cold spectrum is produced by a so-called ‘cold source’, where neutrons pass through and loose energy by inelastic scattering. The spectrum is shifted to longer wavelengths, and owing to larger angles of total reflection, cold neutrons can easily be transported by so-called curved neutron guides to experimental sites. The spectrum is cut from shorter wavelengths and gamma background, both of which are absorbed by shielding material that surrounds the guide. Neutron radiography and tomography, which use the whole reactor spectrum, profit by a very intense neutron beam and utilize the integrated absorption behaviour of the sample; however, some optical problems (wavelength dependency of the angle of total reflection and cold spectrum) of stemming arise from the curvature of the guide, which must be considered in neutron imaging.

## Geometries Used in Radiography and Tomography

Investigating matter by means of neutron radiography and tomography, radiation is usually collimated by slits or a collimator system that allows for a unique correlation between a structure in a sample and its image, which determines spatial resolution. The available intensity limits the spatial resolution, which is shown below. The spatial resolution can be defined as the nearest distance between two points that can be distinguished with at least 10% contrast  $C$  (see also modulation transfer function MTF).  $C$  is defined as

$$C = \left| \frac{g_1 - g_2}{g_1 + g_2} \right|, \quad (18)$$

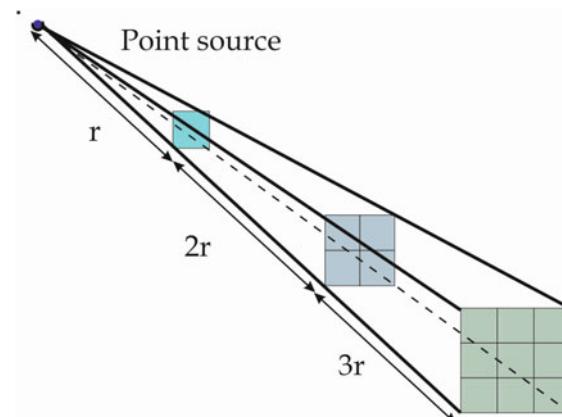
where  $g_{1,2}$  are the gray values in point (1) and (2). In order to detect inhomogeneities, defects, etc., and to distinguish them from each other in matter, one must be aware of some fundamental conditions concerning beam divergence and spatial resolution that also depend on the available intensity (neutrons/[unit area and time]).

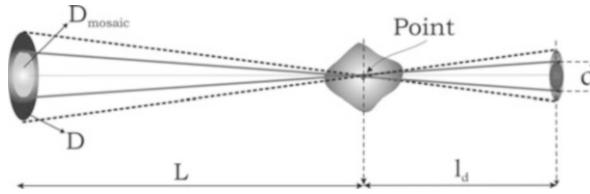
To estimate the intensity in front of a sample, which is placed at a distance  $r$  from a point source, one can deduce the flux as follows. If a source emits radiation in  $4\pi$  the source flux  $I_0$  decreases at a distance as  $1/r^2$  (see Fig. 1). Within a solid angle  $\Delta\Omega = \phi \times \phi$ , the intensity  $I$  is proportional to the area of the surface of the sphere  $= ds \times ds$ ,  $ds \sim \phi \times r$  and  $\phi$  is the one-dimensional divergence of the beam.

Thus, for a divergence  $\phi$  (which is known or can be measured, see below) the available intensity  $I$  in front of a sample can be calculated from a given source flux  $I_0$ . This estimation holds only for an ideal point, and it does not if the point source is used in a neutron guide or in the case of a crystal monochromator. In this case, its integrated reflection power has to be determined, which is proportional to the mosaic spread of the monochromator. The simplest and best way is to measure the neutron intensity for a given cross section (e.g.  $1 \text{ mm}^2$ ) at different distances from the monochromator and plot intensity as a function of  $r$  (distance from monochromator). Using crystal monochromator as source has several advantages, as shown in (Treimer et al. 2010). Therefore, the interpretation and explanation of images depend on, among others, the spatial resolution of the imaging system.

The beam divergence  $\phi$  is very often expressed by the so-called  $L/D$  (read  $L$  over  $D$  ratio). The inverse  $D/L = \phi$ , is the collimation of the beam. However, there are different definitions of the  $L/D$  ratio (see Brenizer 1992; ASTM 1996), but, for simplification, we use the one depicted in Fig. 2. Depending on the size of the source  $D$ , or on the mosaic spread of a crystal source  $D_{\text{mosaic}}$ , a point in matter is enlarged on a screen as  $d = l_d D/L$  (Fig. 2). The definition of the  $L/D$  ratio used in this chapter is as follows:  $D$  = size of the (effective) source,  $L$  = distance source – object (point),  $l_d$  = distance object – detector,  $L/D$  corresponds to the incident beam divergence, i.e.  $D/L \sim \phi$ . A large  $L/D$  means a low divergence and neutron flux and vice versa. In radiography and tomography, one uses both a pin-hole as source and/or a beam

**Fig. 1** The number of neutrons/unit area decreases as  $1/r^2$  ( $r$  = distance from the point sources)





**Fig. 2**  $L/D$  definition used in this chapter. The image of a point in a sample is enlarged to  $d$  in the detector plane. Depending on the size (and structure) of the point source, one gets different enlargements.  $D_{\text{mosaic}}$  signifies a monochromator crystal having a certain mosaic spread

collimation. In the case of a pin-hole,  $D$  is the diameter of the pin-hole. The use of a crystal monochromator in neutron imaging (Herzig 1997; Treimer et al. 2010) involves different geometric conditions for  $L/D$ , which are not discussed here. Investigating, for example, defects of a certain size to be detected in a sample,  $L/D$  must be adapted to the boundary conditions, such as neutron flux, spatial resolution of the detector and possible geometric restrictions (large cryostat or oven). Enlargement of an image of a structure/object (ideally realized with a point source) involves blurring due to the finite size of source and causes smearing of details. Blurring can be described by functions derived in the section “[Resolution Functions \(Gauss, LSF, MTF\)](#).”

To take this blurring into account one has to consider under which conditions (beam divergence) two points in a sample can be uniquely distinguished. The geometry is shown in Fig. 3. For a given  $L/D$ , one can still distinguish two points in sample for a detector distance  $l_d$ , despite the fact that images overlap. The distinctiveness is not only a function of  $L/D$  but also of detector resolution (pixel size).

## Resolution Functions (Gauss, LSF, MTF)

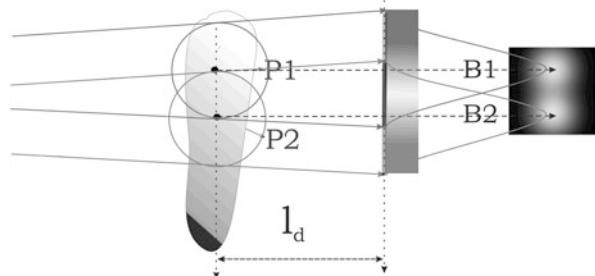
In order to understand the mathematics of spatial resolution and thus blurring of a point onto the detector plane, one considers first a perfect straight edge. Owing to beam divergence  $\phi$ , the image of the straight edge is blurred similar to Fig. 3, as shown for two points. This blurring can (one-dimensionally) be described by the convolution of the step (Heaviside) function with either a Lorentz function or a Gauss function. The Lorentz function, which can be used as a line spread function (LSF), is given by

$$LFS = \frac{\frac{w}{\pi}}{1 + w^2 \cdot (x - x_0)^2} \quad (19)$$

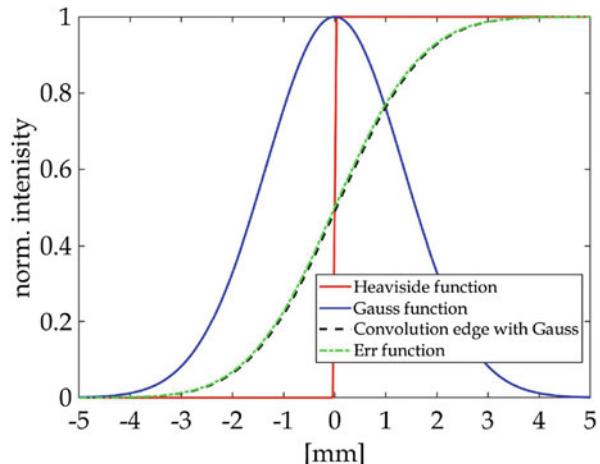
with  $w$  as the full width at half maximum of the function. The Gauss function  $G$  is given by

$$G = \exp\left(-((x - x_0)/\sigma)^2\right). \quad (20)$$

**Fig. 3**  $L/D$  is given, two points in sample are  $l_d$  apart from the detector plane, and despite the fact that images overlap, they can be distinguished from each other



**Fig. 4** Convolution of an Gauss function (blue) with an edge (Heaviside) function (red) simulates the smearing of an image of an edge (black, dotted line), green = error function

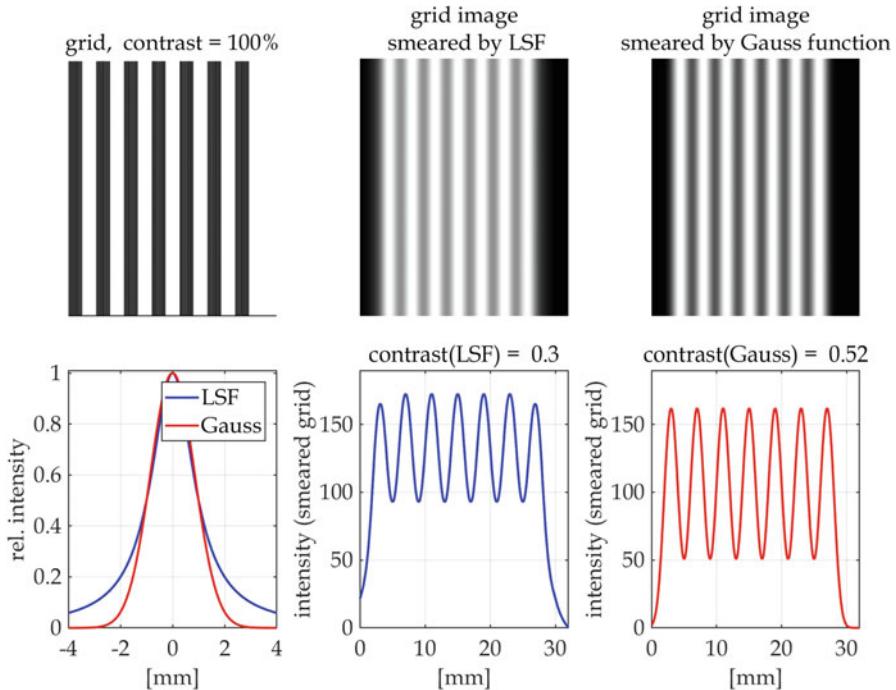


The convolution of a Gauss function (or Lorentz function) with a step function (= Gauss  $\otimes$  Heaviside (step)) simulates the blurring of an image of a sharp edge illuminated by a divergent beam. The convolution Gauss  $\otimes$  Heaviside function is shown in Fig. 4. The shape of the smeared function is similar to an error function, which is the integrated Gauss function, as shown in Fig. 4. Another resolution function is the Lorentz function (Eq. 19), which, when integrated, yields the edge spread function ESF (Eq. 21). The ESF and the error function (Eq. 23) have quite similar shape and describe blurring differently, as shown in Fig. 5.

$$\begin{aligned} \text{ESF}(x) &= \int \frac{w/\pi}{1 + w^2(x - x_0)^2} dx \\ &= \frac{1}{\pi} \arctan(w(x - x_0)) \end{aligned} \quad (21)$$

The integration of the Gauss function (Eq. 20)

$$\int G(x)dx = \int e^{-((x-x_0)/\sigma)^2} dx = F_{\text{err}}(x) \quad (22)$$



**Fig. 5** Smearing of a grid image by a Gauss and a Lorentz function, both having the same FWHM

yields the error function  $F_{\text{err}}(x)$

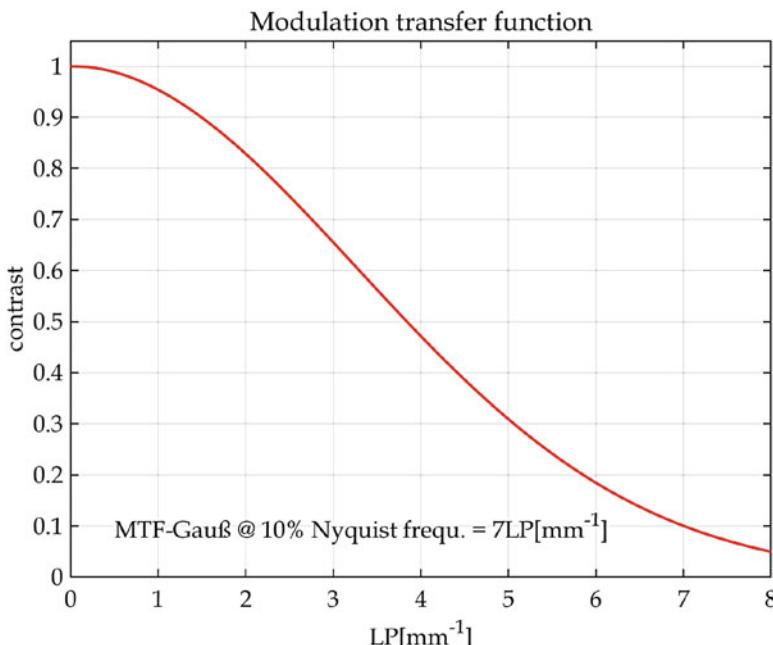
$$F_{\text{err}}(x) = \pi^{(1/2)} \text{erf} \left( (x - x_0) (1/\sigma^2)^{(1/2)} \right) / 2 (1/\sigma^2)^{(1/2)}, \quad \left( \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \right) \quad (23)$$

Applying Gauss function or Lorentz function to a grid, which both have the same FWHM, one recognizes different smearing of the grid, causing different image contrast (blurring), also seen from the line plots. In order to use the correct smearing function for an imaging system, one has to measure the edge blurring by a straight edge and fit the measured function either with Gauss or Lorentz functions; the Pearson function may also be useful.

The spatial resolution of the imaging system can be determined by the so-called modulation transfer function (MTF). It is known as the modulus of the optical transfer function (OTF), which is the Fourier transform (FT) of the ESF. That looks complicated, however it is easy to understand. A finite  $L/D$  ratio is always linked with beam divergence  $\phi$  and a wavelength band  $\Delta\lambda/\lambda > 0$ . Furthermore, a finite precision of scanning and detection (pixel size) also contribute to smearing effects.

If a Gauss or Lorentz function describes a divergent neutron beam, the  $\Delta\lambda/\lambda$  distribution, etc., the convolution with a step function yields a smeared edge function (SEF). With  $w$  (Eq. 19) or  $\sigma$  (Eq. 20) one can tune the shape of the functions and fit (as convolution) measured data. From the fitted shape of the smeared edge by, e.g. a Gauss function, the Fourier transform of the Gauss function,  $\text{FFT}\{G\}$  yields the modulation transfer function MTF; thus, the modulation transfer function (MTF) can be interpreted as a frequency dependent contrast function  $C = C(f)$  (Eq. 18), see Fig. 6. The spatial resolution can be directly read from the 10% contrast; here at 7 LP/mm (appr. 140/ $\mu\text{m}$ ). In this case, 7 LP/mm is called the maximum frequency  $f_{\max}$  that appears in the image with 10% contrast (see Fig. 6). The correct scanning frequency  $f_{\text{scan}}$  (fulfilling the Shannon theorem) is given by  $f_{\text{scan}} = 2 \times f_{\max}$ ; thus, 14 LP/mm (appr. 70  $\mu\text{m}$ ), meaning that the detector pixel size must be 70/ $\mu\text{m}$  or even less to obtain a spatial resolution of 140/ $\mu\text{m}$ .

The modulation transfer function characterizes the spatial resolution of an imaging instrument. It is the image contrast as a function of spatial frequency. The MTF as a frequency spectrum fulfills all features of a Fourier transform, which means that amplitudes as well as rotation are invariants under transformation. Note that the measured MTF is always the product of MTFs of all components. Each component of the system is characterized by its individual  $\text{MTF}_n$ ; thus, the measured MTF is always the product of all  $\text{MTF}_n$ ,  $\text{MTF}_n = \prod_{n=1}^N \text{MTF}_n$ .



**Fig. 6** Modulation transfer function of the image of the grid in Fig. 5

## Mathematics of Computerized Tomography

Computerized tomography, developed in the 1970s, is one of the most powerful nondestructive techniques. The mathematical basis is the so-called Radon transform, published in 1917 (Radon 1917). However, the real importance of this transform was detected by the radio astronomer Ron N. Bracewell (Bracewell 1956), who presented the first solution of the emitted microwave radiation problem. The relation to the Fourier transform (slice theorem) involves an absolutely essential tool for image reconstruction, which was later proved by a number of authors.

The principal problem of CT to be solved is the reconstruction of an object from projections. The mathematics is presented by many authors (Herman 1980; Rosenfeld and Kak 1982); however, here we will follow the book of A. C. Kak and M. Slaney (Parts of this chapter were presented during workshops and lectures on tomography in (Anderson et al. 2009)), (Rosenfeld and Kak 1982; Kak and Slaney 1999) and simplify some mathematics.

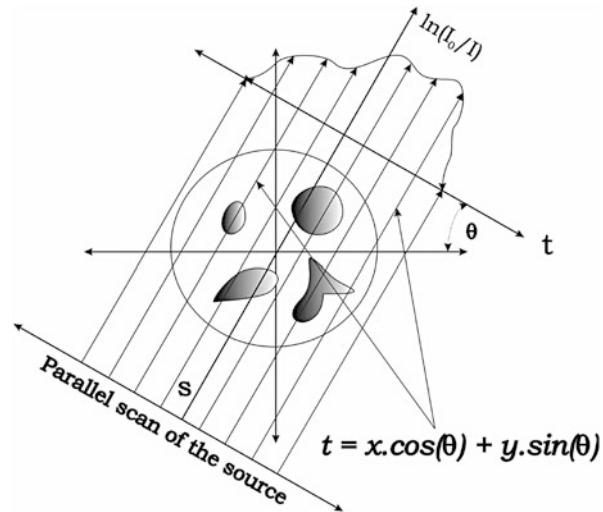
The basic mathematics of computerized tomography (CT) uses the simple case of parallel beam geometry. In practice, ‘fan beam’ or ‘cone beam’ geometries are used, the mathematical transformation from parallel to fan or cone beam geometry is given in the section “[Basics of Neutron Radiography and Tomography](#)” in (Kak and Slaney 1999). The object  $O = O_z(x, y, z)$  under investigation is ‘cut’ in ‘z’ slices  $O_z(x, y)$ . Each slice is scanned by straight lines under angles  $\theta$ ,  $0 \leq \theta \leq \pi$  producing ‘projections’  $P(\theta, t)$ . This mapping, defined by the set (infinite number) of straight lines/projections, is called the (two-dimensional) ‘Radon transform’ of the  $z$ th slice of  $O = O_z(x, y, z), \hat{R}[O_z]$ ,

$$\widehat{O}_z = \hat{R}[O_z] = \int_{\text{path}} O_z(x, y) ds. \quad (24)$$

A slice  $O_z(x, y)$  in the  $\{x, y\}$  coordinate system maybe characterized by a function  $\mu = \mu(x, y)$ , which defines the two-dimensional dependence of physical parameter  $\mu$  in the  $\{x, y\}$  plane. In medicine or nondestructive testing,  $\mu = \mu(x, y)$  is the position dependent radiation attenuation parameter of the object. A ray traversing the slice (Fig. 7) is attenuated as

$$I(t, \theta) = I_0 \cdot \exp \left\{ - \int_{\text{path}} \mu(x, y) \cdot ds \right\} \quad (25)$$

where  $I_0$  is the incident intensity (number of photons, neutrons, etc.),  $\theta$  the angle between  $t$  and  $x$  axis and  $I(t, \theta)$  is the transmitted intensity behind the object. The path (line) integral is calculated by ‘summing up’ the attenuation along this path only at coordinate  $t$ . We rewrite (Eq. 25) and define  $P_\theta(t)$  as follows:

**Fig. 7** Parallel scanning

$$\ln\left(\frac{I_0}{I}\right) = P_\theta(t) = \int_{\text{Path}} \mu(x, y) \cdot ds \quad (26)$$

$P_\theta(t)$  is the set of all path integrals scanned under the angle  $\theta$  and is called ‘projection’ of  $\mu = \mu(x, y)$ . To guarantee that only points are taken into account that lie on the straight line at ‘ $t$ ’,  $t$  (in the rotated  $\{t, s\}$  coordinate system) must be expressed in the  $\{x, y\}$  coordinate system, that means the rotation matrix  $R_z$  must be applied to each pair  $(x, y)$ ,

$$R_z = \begin{Bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{Bmatrix} \Rightarrow \begin{Bmatrix} t \\ s \end{Bmatrix} = R_z \cdot \begin{Bmatrix} x \\ y \end{Bmatrix} \quad (27)$$

Then one gets equations for the parameters  $t$  and  $s$  as functions of the coordinates  $x, y$  and angle  $\theta$

$$\begin{aligned} t &= x \cdot \cos(\theta) + y \cdot \sin(\theta) \rightarrow 0 = t - x \cdot \cos(\theta) - y \cdot \sin(\theta) \\ s &= -x \cdot \sin(\theta) + y \cdot \cos(\theta) \rightarrow 0 = s + x \cdot \sin(\theta) + y \cdot \cos(\theta) \end{aligned} \quad (28)$$

In order to calculate projections, the path integral (Eq. 26) is transformed into an area integral,

$$\int_{\text{path}} \mu(x, y) \cdot ds \rightarrow \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x, y) \cdot dx \cdot dy \quad (29)$$

Only points on the straight line must contribute to one point ‘ $t$ ’ of the projection  $P(\theta, t)$ ; thus, we use the  $\delta$  function

$$\delta(t - t_1) = \begin{cases} 1 & \text{for } t = t_1 \\ 0 & \text{for } t \neq t_1 \end{cases} \quad (30)$$

and express (Eq. 26) as a double integral over the  $\{x, y\}$  plane:

$$P_\theta(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy \mu(x, y) \delta(x \cdot \cos(\theta) + y \cdot \sin(\theta) - t) \quad (31)$$

Note, that one single point  $t_1$  of the projection  $P = P_\theta(t_1)$  must be calculated by the double integral

$$P_\theta(t_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy \mu(x, y) \delta(x \cdot \cos(\theta) + y \cdot \sin(\theta) - t_1)$$

The scanning procedure is as follows: All points of  $P = P_\theta(t_1)$  are recorded under a given angle  $\theta$  by mapping the two-dimensional function  $\mu(x, y)$  to a projection; the set of all projections is then the Radon transform of  $\mu(x, y)$ . As stated above, the goal of CT is to reconstruct  $\mu(x, y)$  from projections.

There are several methods to reconstruct  $\mu(x, y)$  known as inversion methods of the Radon transform. Besides algebraic procedures, such as Algebraic Reconstruction Technique (ART) or Simultaneous Iterative Reconstruction Technique (SIRT) (Kak and Slaney 1999; Andersen and Kak 1984), Fourier methods are the most powerful ones. In order to understand – in principle – the Fourier method, one must understand the underlying mathematics based on the so-called ‘projection-slice theorem’.

## Projection-Slice Theorem

The projection-slice theorem, shortly slice-theorem, provides a connection between the two-dimensional Fourier transform of  $\mu(x, y)$  and the one-dimensional Fourier transform of the Radon transform of  $\mu(x, y)$  (Deans 1983):

**The projection-slice-theorem states, that if  $P_\theta(t)$  has a Fourier Transform  $F\{P_\theta(t)\}$  and  $\mu(x, y)$  has a Fourier Transform  $\Gamma\{\mu(x, y)\}$ , then  $F\{P_\theta(t)\} \subset F\{\mu(x, y)\}$ .**

We prove first a simple case ( $\theta = 0$ ) to understand the idea of the theorem and assume that  $\mu(x, y)$  is finite and has a Fourier transform  $M(u, v)$  as (This part closely follows section “Resolution Functions (Gauss, LSF, MTF)” in (Kak and Slaney 1999)).

$$\begin{aligned} \mu(x, y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} M(u, v) e^{j2\pi(u.x+v.y)} du dv \\ M(u, v) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x, y) e^{-j2\pi(u.x+v.y)} dx dy \end{aligned} \quad (32)$$

The projection  $P_\theta(t)$  shall have a Fourier transform  $\mathcal{F}\{P_\theta(t)\}$  as

$$S_\theta(\omega) = \int_{-\infty}^{\infty} P_\theta(t) \cdot e^{-i \cdot 2\pi \omega t} dt \quad (33)$$

$\theta = 0$  means that the coordinate systems  $\{x, y\}$  and the  $\{t, s\}$  are congruent, i.e. their scales are the same. The Fourier transform  $M(u, v)$  along the straight line  $v = 0$  in the frequency domain reduces  $M(u, v)$  (Eq. 32) to

$$M(u, 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x, y) \cdot e^{-i \cdot 2\pi(u \cdot x)} dx \cdot dy \quad (34)$$

The exponential function depends on  $x$ , only; therefore, one can write

$$M(u, 0) = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} \mu(x, y) \cdot dy \right] \cdot e^{-i \cdot 2\pi u \cdot x} dx \quad (35)$$

From (Eq. 26) one can set

$$\begin{aligned} P_\theta(t) &= \underset{\text{Path}}{\int} \mu(x, y) \cdot ds \rightarrow (t \sim x \text{ und } s \sim y) \\ P_{\theta=0}(x) &= \int_{-\infty}^{\infty} \mu(x, y) \cdot dy \end{aligned} \quad (36)$$

$P_{\theta=0}(x)$  is a projection of  $\mu(x, y)$  at  $\theta = 0$  along lines with  $x = \text{const}$ . Replacing  $\mu(x, y)$  by  $P_{\theta=0}(x)$  in (Eq. 32) one gets

$$M(u, 0) = \underbrace{\int_{-\infty}^{\infty} P_{\theta=0}(x) \cdot e^{-i \cdot 2\pi \cdot u \cdot x} dx}_{=\text{Fourier Transform of } P_{\theta=0}(t)} \quad (37)$$

Thus, the Fourier transform of the projection  $P_\theta(t)$  agrees with the two-dimensional Fourier transform of  $\mu(x, y)$  for  $\theta = 0$ , corresponding  $v = 0$ :

$$M(u, 0) = S_{\theta=0}(u) \quad (38)$$

This is the simplest form of the Slice Theorem which can be generalized as follows:

If  $\mu(x, y)$  has a Fourier transform  $M(u, v)$  and if  $S(\omega)$  is the Fourier transform of  $P_\theta(t)$ , then  $S(\omega) \subset M(u, v)$ .

This important theorem can be proved as follows. Consider a projection  $P_\theta(t)$  in the rotated coordinate system  $\{t, s\}$ . The rotated coordinates  $\{t, s\}$  are obtained applying the rotation matrix on  $\{x, y\}$ ,

$$\begin{bmatrix} t \\ s \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (39)$$

then  $P_\theta(t)$  in (Eq. 26) can be written as ( $\mu(x, y) \rightarrow \mu(t, s)$ )

$$P_\theta(t) = \int_{-\infty}^{\infty} \mu(t, s) \cdot ds \quad (40)$$

Note that  $P_\theta(t)$  is (still) a line integral along lines of constant  $t$ .  $P_\theta(t)$  has a Fourier transform  $S_\theta(\omega)$  (s. (Eq. 33)). Substituting  $P_\theta(t)$  by  $\mu(x, y)$ ,  $S_\theta(\omega)$  becomes

$$S_\theta(\omega) = \int_{-\infty}^{\infty} P_\theta(t) \cdot e^{-i \cdot 2\pi \cdot \omega \cdot t} dt = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} \mu(t, s) \cdot ds \right\} \cdot e^{-i \cdot 2\pi \cdot \omega \cdot t} dt \quad (41)$$

The change from the  $\{x, y\}$  system to  $\{t, s\}$  system and vice versa is given by (Eq. 27)

$$\begin{bmatrix} t \\ s \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Thus, the parameter  $t$  is  $t = x \cdot \cos(\theta) + y \cdot \sin(\theta)$ , known from (Eq. 28); so,  $S_\theta(\omega)$  in (Eq. 41) can be written in the  $\{x, y\}$  system as

$$S_\theta(\omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x, y) \cdot e^{-i \cdot 2\pi \cdot \omega \cdot (x \cdot \cos(\theta) + y \cdot \sin(\theta))} dx \cdot dy \quad (42)$$

The correlation of the polar coordinates  $(\omega, \theta)$  to  $(x, y)$  is shown in Fig. 8, and with  $u = x \cdot \cos(\theta)$  and  $v = y \cdot \sin(\theta)$ ,  $S_\theta(\omega)$  becomes

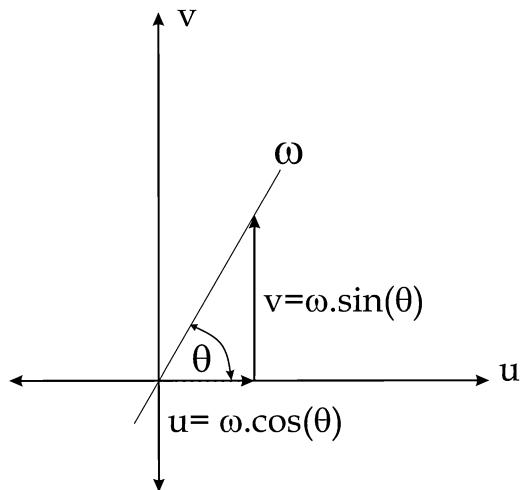
$$S_\theta(\omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x, y) \cdot e^{-i \cdot 2\pi \cdot (u \cdot x + v \cdot y)} dx \cdot dy = M(u, v) \quad (43)$$

The Fourier transform of  $P_\theta(t) = S_\theta(\omega)$  becomes

$$S_\theta(\omega) = M(u, v) = M(\omega \cdot \cos(\theta), \omega \cdot \sin(\theta)) = M(\omega, \theta) \quad (44)$$

The double integral is exactly the Fourier transform of  $\mu(x, y)$  (s. (Eq. 32)); however, in polar coordinates. The left hand side of Eq. (44) is the Fourier transform of the projection  $P_\theta(t)$ , the right hand side is the Fourier transform of  $\mu(x, y)$ . Therefore, Fourier-slice theorem can be formulated as follows:

**Fig. 8** Correlation ( $u, v$ ) coordinates to polar coordinates ( $\omega, \theta$ )



**The Fourier Transform of projection  $P_\theta(t)$ ,  $\mathcal{F}\{P_{\text{theta}}\}(t)$  is found as part of the Fourier Transform of  $\mu(x, y)$ , however in polar coordinates, because the rotation is invariant under the Fourier Transform.**

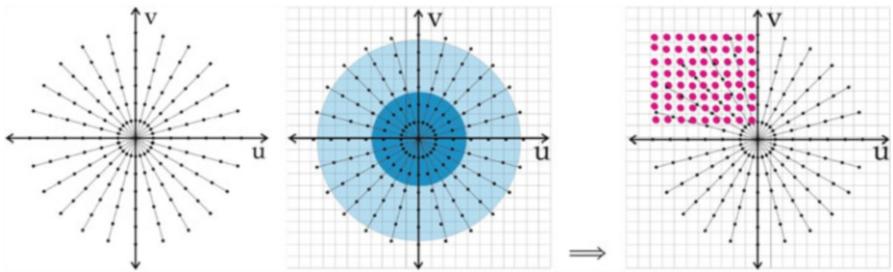
Therefore, one has a tool to ‘fill up’ the Fourier space with projections (i.e. with their Fourier transform) and get  $M(u, v)$  for a infinite number of scans. In that case, one gets  $\text{FT}\{\mu(x, y)\} = M(u, v)$  and with the Fourier transform of  $M(u, v)$ ,  $\text{FT}\{M(u, v)\} = \mu(x, y)$

$$\mu(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} M(u, v) \cdot e^{+i.2\pi(u.x+v.y)} du.dv \quad (45)$$

In reality one has a limited number of projections that have to fit to the rectangular coordinate system, and obviously, lower frequencies are much better defined than higher ones because interpolation for high frequency data is less well defined than for low frequency data (Fig. 9). That means that details, edges, small structures, etc., in real space appear unshapely and blurred.

That can be improved by increasing the number of projections, but there is a (mathematical) limit that restricts this number, which means that increasing the number of projections does not improve the appearance of the reconstruction. This can be shown by applying the ‘Nyquist-Shannon theorem’ to CT.

In digital recording the effective pixel size of a detector is a measure of spatial resolution. In order to get a number for a correct sampling of a projection, one considers a projection, e.g. as shown in Fig. 7. Assume a sampling distance  $T$  (= pixel size of a 1D detector) and  $T$  fulfills the Nyquist-Shannon-Sampling theorem (or WKS-sampling theorem, WKS = Whittaker, Kotel-nikow and Shannon) as follows:



**Fig. 9** From left to right: Frequency domain, lines are FT of  $P$ , low frequencies are ‘more dense than higher ones’,  $\Rightarrow$  data fitted to rectangular coordinates,  $(\omega, \theta) \rightarrow (u, v)$

A continuous (limited) function  $h = h(x)$  can uniquely be reconstructed from its discrete data set  $h[x] = h[n \cdot T]$  ( $n \in \mathbb{N}$ ), if  $h(x)$  is sampled with twice the highest frequency  $f_{\max}$  of its Fourier spectrum  $H(f)$ . Thus, the sampling frequency  $f_T = 2f_{\max}$ .

One must add a given contrast, usually between 10% and 20% (depending on the imaging system); in Fig. 6 the limiting contrast was 10%, defining  $f_{\max} = 7 \text{ LP/mm}$ , i.e. the sampling frequency  $f_T$  in Fig. 6 becomes 14 LP/mm corresponding to sampling distance  $T = 1/f_T = 0.071 \text{ mm}$ .

The number of projections for a unique 2D slice reconstruction depends on the number of sampling points per projections. If an object has a maximum diameter  $d$ , and  $T$  fulfills the condition of the Nyquist-Shannon-Sampling theorem, then  $d$  must be sampled  $T \times N$  and  $N$  is the number of sampling points ( $N = d/T$ ). Furthermore, one requires that the sampling of the half perimeter ( $d \cdot \pi/2$ ) of the circle (s. Fig. 10) also fulfills the Nyquist-Shannon-Sampling theorem with the same sampling distance  $T$ . Therefore, if  $M$  is the number of sampling points of the half perimeter, then  $M \cdot T = d \cdot \pi/2 = T \cdot N \cdot \pi/2$ . Using the same  $T$  for both samplings, one gets

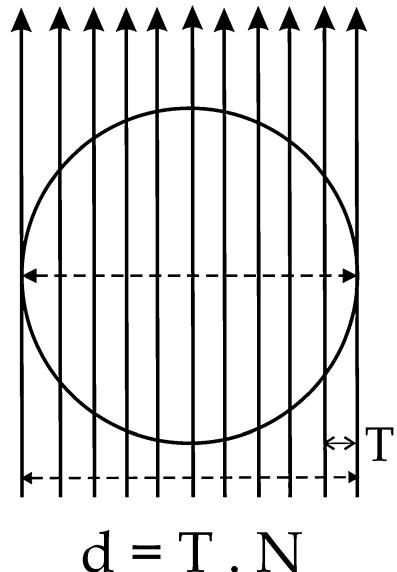
$$M = \frac{\pi}{2} N \leftrightarrow d = \frac{2}{\pi} \cdot T \cdot M \quad (46)$$

Note that the correctly reconstructed slice diameter depends on both  $N$  = number of sampling points (detectors) of one projection and  $M$  = number of projections. Decreasing  $M$  decreases the diameter of the correctly reconstructed slice; however, the reconstructed slice will still have the diameter  $d$  (see Eq. 46).

## Theory of Back-Projection

The underlying problem in CT is the reconstruction of a 2D function from projections; strictly speaking, the problem to be solved is the mathematical derivation of the reconstruction of a slice  $\mu(x, y)$  from projections. The data acquisition is performed in a polar coordinate system because either the sample rotates or the source and detector system rotate around the sample (except for medical applications, there are only a few

**Fig. 10** Shannon theorem applied to scans in tomography, see text



$$d = T \cdot N$$

examples for the latter geometry), and this holds for all scan options (parallel-, fan- and cone-beam geometry). Again we follow the derivation given in (Kak and Slaney 1999). The Fourier transform of  $\mu(x, y)$  is given by (Eq. 32),

$$\mu(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} M(u, v) \cdot e^{+i.2\pi(u.x+v.y)} du.dv \quad (47)$$

Both sides are represented in a rectangular coordinate system,  $\{(x, y)\}$  and  $\{(u, v)\}$ . The scanning happens with polar coordinates, and as the slice-projection theorem states that the FT of  $P_\theta(t)$  is part of  $M(u.v)$  (Eq. 38), one can step by step transform the double integral into polar coordinates. One has to replace  $u, v, du$  and  $dv$  by  $\theta, \omega, d\theta$  and  $d\omega$ . From Fig. 8 one gets

$$u = \omega \cdot \cos(\theta) \quad v = \omega \cdot \sin(\theta) \quad (48)$$

In order to transform a function  $F(u, v)$  to  $F(\theta, \omega)$ , the double integral (Eq. 38) must be rewritten as

$$\iint F(u, v).du.dv = \iint F(\omega, \theta)d\omega.d\theta. \mid \text{Jacobydet} \mid \quad (49)$$

First of all, one must determine the Jacobi-Determinate:

$$\left| \begin{array}{c} \frac{\delta u}{\delta \omega} \frac{\delta u}{\delta \theta} \\ \frac{\delta v}{\delta \omega} \frac{\delta v}{\delta \theta} \end{array} \right| = \left| \begin{array}{c} \cos(\theta) - \omega \cdot \sin(\theta) \\ \sin(\theta) \quad \omega \cdot \cos(\theta) \end{array} \right| = \omega (\cos^2(\theta) + \sin^2(\theta)) = \omega, \quad (50)$$

then the integration limits must be changed to  $\omega = 0 \dots \infty$  and  $\theta = 0 \dots 2\pi$ . With (Eq. 47), (Eq. 49) and (Eq. 50) one gets

$$\mu(x, y) = \int_0^{2\pi} \int_0^\infty M(\omega, \theta) \cdot e^{i2\pi\omega \cdot (x \cdot \cos(\theta) + y \cdot \sin(\theta))} \omega \cdot d\omega \cdot d\theta \quad (51)$$

Now the integral (Eq. 51) is split into two parts:  $\theta = 0 \dots 180^\circ$  and  $\theta = 180^\circ \dots 360^\circ$ :

$$\begin{aligned} \mu(x, y) &= \underbrace{\int_0^{\pi/2} \int_0^\infty M(\omega, \theta) \cdot e^{i2\pi\omega \cdot (x \cdot \cos(\theta) + y \cdot \sin(\theta))} \omega \cdot d\omega \cdot d\theta}_{(I)} \\ &\quad + \underbrace{\int_{\pi/2}^{\pi} \int_0^\infty M(\omega, \theta + 180^\circ) \cdot e^{i2\pi\omega \cdot (x \cos(\theta + 180^\circ) + y \sin(\theta + 180^\circ))} \omega \cdot d\omega \cdot d\theta}_{(II)} \end{aligned} \quad (52)$$

From the scanning geometry (Fig. 11), one sees that  $M(\theta + 180^\circ) = M(-\omega, \theta)$ . So, the second term (II) in (Eq. 49) can be written as

$$\begin{aligned} (II) &\rightarrow \int_0^{\pi/2} \int_0^\infty M(-\omega, \theta) \cdot e^{i2\pi(-\omega) \cdot (x \cdot \cos(\theta + 180^\circ) + y \cdot \sin(\theta + 180^\circ))} (\omega) \cdot d\omega \cdot d\theta \\ &\rightarrow \int_0^{\pi/2} \int_0^\infty M(-\omega, \theta) \cdot e^{i2\pi(-\omega) \cdot \underbrace{(x \cdot \cos(\theta + 180^\circ) + y \cdot \sin(\theta + 180^\circ))}_{-x \cdot \cos(\theta) + y \cdot \sin(\theta)}} (\omega) \cdot d\omega \cdot d\theta \end{aligned}$$

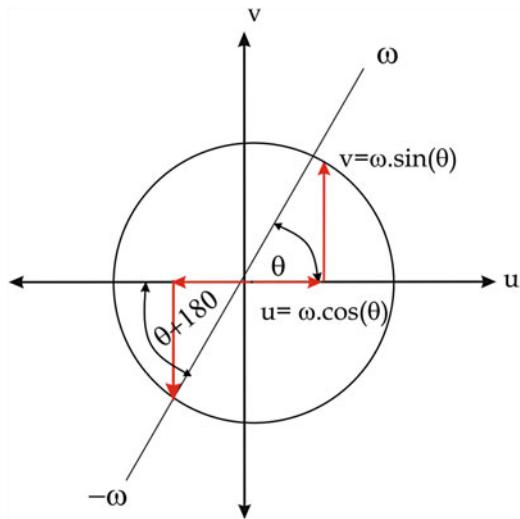
using  $\cos(\theta + 180^\circ) = -\cos(\theta)$  and  $\sin(\theta + 180^\circ) = -\sin(\theta)$  one gets

$$\rightarrow \int_0^{\pi/2} \int_0^\infty M(-\omega, \theta) \cdot e^{i2\pi(-\omega) \cdot (x \cdot \cos(\theta) + y \cdot \sin(\theta))} (-\omega) \cdot d\omega \cdot d\theta$$

inverting the  $\omega$  integration yields

$$\rightarrow \int_{-\infty}^{\pi/2} \int_0^\infty M(\omega, \theta) \cdot e^{i2\pi\omega \cdot (x \cdot \cos(\theta) + y \cdot \sin(\theta))} (-\omega) \cdot d\omega \cdot d\theta$$

**Fig. 11** Scanning geometry  
rotation angle  $\theta$  and  $\theta + 180^\circ$



Therefore, (Eq. 52) becomes

$$\begin{aligned} \mu(x, y) &= \int_0^{\pi} \int_0^{\infty} M(\omega, \theta) \cdot e^{+i2\pi\omega(x \cos(\theta) + y \sin(\theta))} \omega \cdot d\omega \cdot d\theta \\ &+ \int_0^{\pi} \int_0^{\infty} M(\omega, \theta) \cdot e^{i2\pi\omega(x \cos(\theta) + y \sin(\theta))} (-\omega) \cdot d\omega \cdot d\theta \end{aligned}$$

which can be simplified to

$$\mu(x, y) = \int_0^{\pi} \int_{-\infty}^{\infty} M(\omega, \theta) \cdot e^{i2\pi\omega(x \cos(\theta) + y \sin(\theta))} |\omega| \cdot d\omega \cdot d\theta \quad (53)$$

Now one uses the relation (see Eq. 48)

$$u = \omega \cdot \cos(\theta) \quad v = \omega \cdot \sin(\theta) \quad (54)$$

and writes the Fourier transform of  $\mu(x, y)$  in polar coordinates as

$$\mu(x, y) = \int_0^{\pi} \int_{-\infty}^{\infty} M(\omega, \theta) \cdot e^{i2\pi\omega \cdot t} |\omega| \cdot d\omega \cdot d\theta \quad (55)$$

The Fourier transform of  $\mu(x, y)$  (Eq. 55) expressed in polar coordinates contains the Fourier transform of  $P(t, \theta) = S(\omega, \theta)$ , (Eq. 33), multiplied by  $|\omega|$ ,

$$\mu(x, y) = \int_{0-}^{\pi} \int_{-\infty}^{\infty} S(\omega, \theta) \cdot |\omega| \cdot e^{i2\pi\omega t} \cdot d\omega \cdot d\theta \quad (56)$$

If one considers first the integral over  $\omega$  in (Eq. 56),

$$\int_0^{\pi} \int_{-\infty}^{\infty} \underbrace{S(\omega, \theta) \cdot |\omega|}_{\text{Product}} \cdot e^{i2\pi\omega t} \cdot d\omega \cdot d\theta \\ Q_\theta(t) \sim \text{FourierTransform}$$

one finds an  $\omega$ -dependent product function as a Fourier transform of a function  $Q(t)$ . Setting

$$Q_\theta(t) = \int_{-\infty}^{\infty} S(\omega, \theta) \cdot |\omega| \cdot e^{i2\pi\omega t} \cdot d\omega, \quad (57)$$

(Equation 57) represents a Fourier transform of the product  $S(\omega, \theta) \cdot |\omega|$ . That means (in the frequency domain)  $|\omega|$  acts as a filter function on  $S(\omega, \theta)$ . The second integration of  $\mu(x, y)$  becomes  $t_\theta = x \cdot \cos(\theta) + y \cdot \sin(\theta)$  (see Eq. 28)

$$\mu(x, y) = \int_0^{\pi} Q_\theta(t) \cdot d\theta = \int_0^{\pi} Q_\theta(x \cdot \cos(\theta) + y \cdot \sin(\theta)) \cdot d\theta \quad (58)$$

$Q_\theta(t)$  is called the filtered projection. (Eq. 58) requires a special explanation. The values of  $\mu(x, y)$  are given by the line integrals along the paths  $t = x \cdot \cos(\theta) + y \cdot \sin(\theta)$  summed up for scanning angles  $0 = 0 \dots 180^\circ$ . That means that every point  $(x, y)$  in  $\mu(x, y)$  is the result of the (back)projected values of the filtered projection  $Q_\theta$ . The filtering process must be performed before the back projection. Recalling that the Fourier transform of a product of two functions is the convolution of the Fourier transforms of these functions:

$$\begin{aligned} f(x) \otimes g(x) &= \int_{-\infty}^{\infty} f(t) \cdot g(x-t) dt \quad (\text{convolution}) \\ FT\{f(x) \otimes g(x)\} &= F(\kappa) \cdot G(\kappa) \quad (\text{Fourier transform of convolution}) \Rightarrow \\ FT\{F(\kappa) \cdot G(\kappa)\} &= \int_{-\infty}^{\infty} F(\kappa) \cdot G(\kappa) \cdot e^{i2\pi x \cdot \kappa} d\kappa \Rightarrow \\ &= f(x) \otimes g(x) \end{aligned} \quad (59)$$

Therefore, one gets an easy procedure for the reconstruction:

1. collect projections  $P_\theta(t)$ ,  $\theta = 0 \dots 180^\circ$
2. convolution  $P_\theta(t)$  with Filter kernel  $p(t)$ ; ( $p(t) \sim \mathcal{FT}\{|\omega|\}$ , see below)
3. back projection of the convolution  $P_\theta(t) \otimes p(t)$

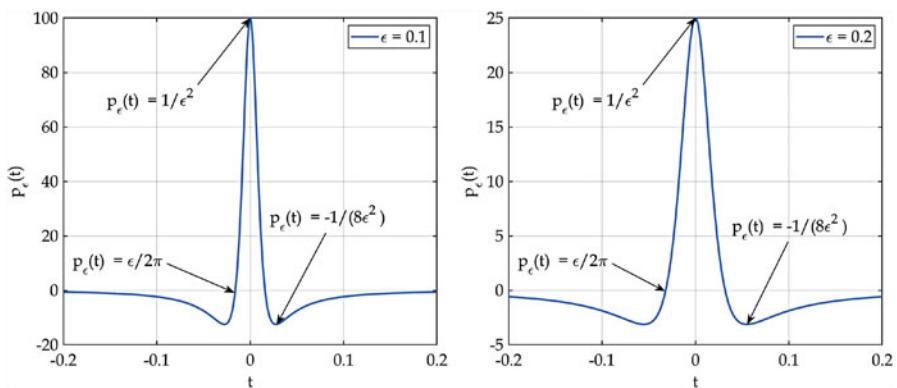
For the filtering process, the Fourier transform of  $|\omega|$ ,  $\mathcal{FT}\{|\omega|\}$  must be known, in order to express (Eq. 57)

$$Q_\theta(t) = \int_{-\infty}^{\infty} P_\theta(\alpha) \cdot p(t - \alpha) d\alpha \quad (60)$$

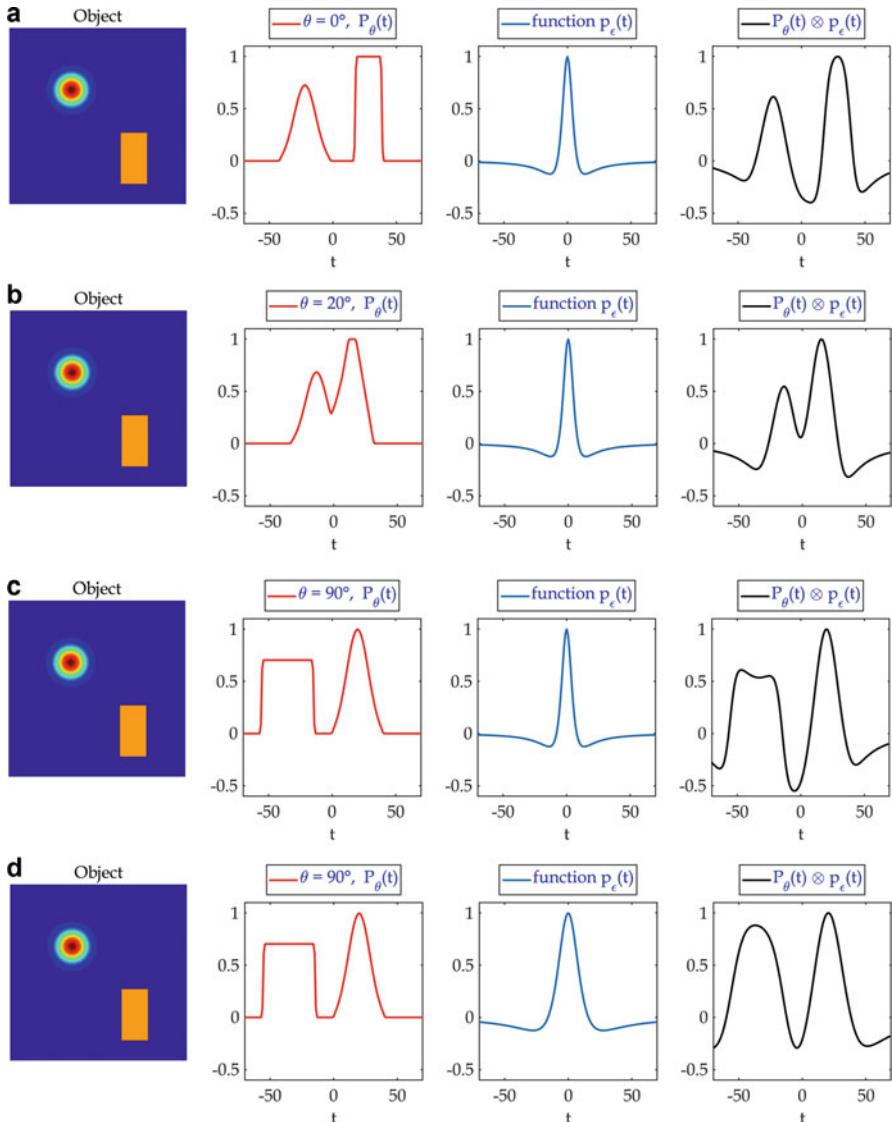
Unfortunately,  $\mathcal{FT}\{|\omega|\}$  does not exist; however, one can approximate  $|\omega|$  by  $|\omega| \exp[-\epsilon|\omega|]$ , as  $\epsilon \rightarrow 0$ . The  $\mathcal{FT}\{|\omega| \exp[-\epsilon|\omega|]\} = p_\epsilon(t)$  exists as

$$p_\epsilon(t) = \frac{\epsilon^2 - [2\pi \cdot t]^2}{(\epsilon^2 + [2\pi \cdot t]^2)^2} \quad (61)$$

The function  $p_\epsilon(t)$  (Eq. 61) depends on the characteristic parameter  $\epsilon$ , which influences the shape of  $p(t)$  as shown in Fig. 12. Thus, the important parameter for the convolution with  $P_\theta(t)$  is 6, which tunes the convolution (Eq. 60, Fig. 13). Besides different  $\epsilon$  values in  $p_\theta(t)$ , the number of projections is important for the quality of a reconstruction of an object (see Eq. 46), which is demonstrated in Fig. 14. To check ‘M’, the correct number of projections, the objects were placed in a field of  $141 \times 141$  pixels, the length of a projection is the diagonal of the square, i.e.  $\sqrt{2} \times 141$  corresponding to 200 pixel; thus, the number of necessary projection

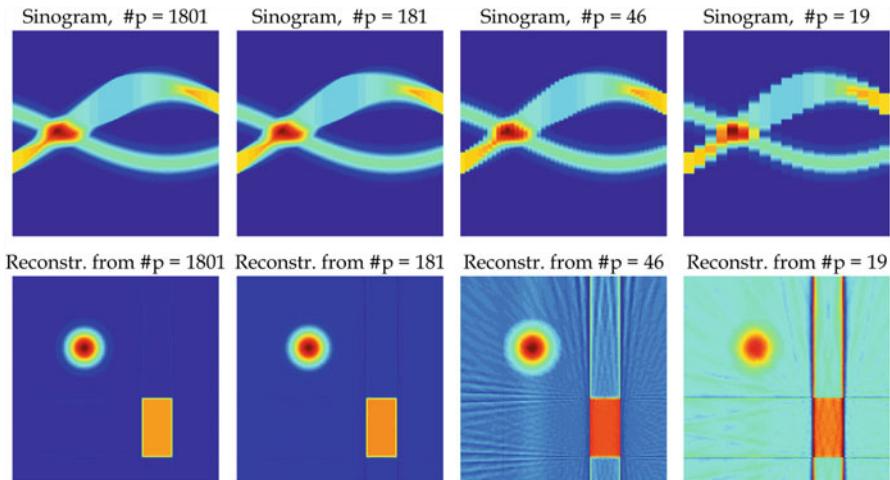


**Fig. 12** Convolution-kernel  $p_\epsilon(t)$  for  $\epsilon = 0.1$  (left graph) and  $\epsilon = 0.2$  (right graph). Note, the maximum is always  $1/\epsilon^2$ , minimum  $= -1/8\epsilon^2 p_\epsilon(t) = 0$  for  $\epsilon/2\pi$

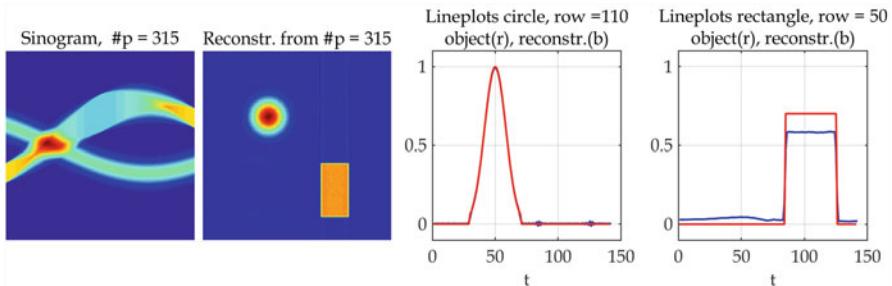


**Fig. 13** Image and plots in (a–e) show different projections ( $\theta = 0^\circ$ ,  $20^\circ$  and  $90^\circ$ ),  $P_\theta(t)$  with  $\epsilon = 0.1$ , (d) is the same projection as (c), but  $P_\theta(t)$  with  $\epsilon = 0.2$ . Note the different convolution product of (c) and (d)

$M = 315$  (cp. (Eq. 46)). The reconstruction is shown in Fig. 15 and a comparison with the reconstruction from 1801 projections in Fig. 14 shows no difference, which means that the nearly six times ‘oversampling’ associated with enormous data excess can be avoided.



**Fig. 14** Sinograms (upper row) and reconstructions from different number of projections  $\#p$



**Fig. 15** Comparison of original object from Fig. 13 and reconstruction from Fig. 14

---

## Fundamentals of Radiography and Tomography with Polarized Neutrons

### Some Historical Remarks

The study of magnetism in the bulk of matter is of principal interest in materials science and also in fundamental physics. This topic found its way very early into classical textbooks of neutron physics (Bacon 1975). There are only a few nondestructive methods that can look below the surface inside of samples having magnetic structures or fields. A suitable way for doing that is to use the magnetic moment of polarized neutrons to investigate these magnetic structures in samples. The advantages of utilizing neutrons are low attenuation by matter and no charge;

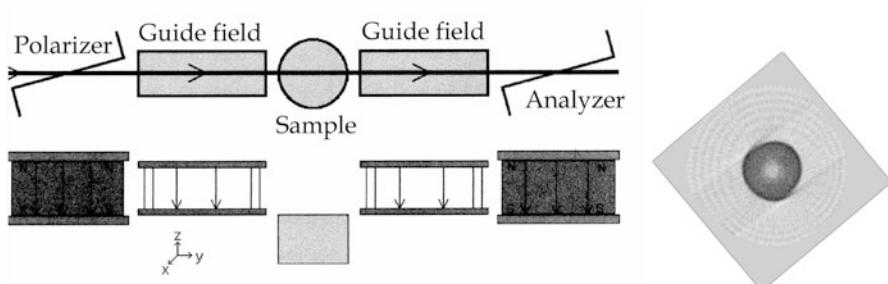
both of which allow for detailed studies of magnetism in the interior of samples by means of polarized neutron radiography and tomography.

The study of magnetic structures in matter is based on the neutron depolarization concept introduced in 1941 by Halpern and Holstein (Halpern and Johnson 1939). They examined the transmission of polarized neutrons through magnetic domains that are regularly or irregularly magnetized, and in 1973 Rekveldt expanded this concept for three dimensional techniques describing the change of polarization by a  $[3 \times 3]$  depolarization matrix (Rekveldt 1973, 1976). This idea for neutron radiography and tomography was mainly further developed by (Hochhold et al. 1996; Badurek et al. 1997a, b, 1998; Leeb et al. 1998, 2005; Jericha et al. 2007). Further theoretical considerations can be found in (Hilger 2010).

A different method of reconstructing magnetic fields in matter uses the Radon transform of the magnetic field, which reduces the spin-interaction with the magnetic field to the calculation of the number of Larmor precessions of the neutron spin, only taking into account reconstructing the amount of the magnetic field (Treimer et al. 2012a, b).

However, the very first experiment concerning polarized neutron imaging was done independently of the publications mentioned above and was performed in 1996/1997. The experiment was carried out by the ‘E8’ instrument at the Berlin research reactor (BER II) and the first 2D reconstruction of a stray field of a small magnet from only a few projections could be realized because the stray field was assumed to be radially symmetric (see Fig. 16) (Kraft 1996; Schaper 1996; Ernst 1997; Herzig 1997).

The set up was simple (see Fig. 16). A perfect Germanium single crystal served as monochromator, and neutrons were polarized by a supermirror. Behind a guide field, a small permanent magnet (diameter = 20 mm with hole = 5 mm) was translated approximately 60 mm perpendicular to the neutron beam, which was collimated by two slits (width = 0.5 mm) in front and behind the sample, defining the spatial resolution for the reconstruction. The spin was analysed downstream by a second supermirror, i.e. neutrons were totally reflected if their spin were parallel to the magnetization of the supermirror. In order to get rid of the high background in the



**Fig. 16** Part of the ‘E8’ instrument at the BER II, the first set up for imaging with polarized neutrons (1996/1997). Right image is the first 2D reconstruction of the magnetic stray field of a small permanent magnet, see text

direct beam, neutrons were Bragg reflected by a graphite crystal towards a He<sup>3</sup> detector (not shown in Fig. 16 as well as the Ge monochromator in the neutron guide). The oscillating structure of the 2D reconstruction of the stray fields obtained from neutron raw data proved the spatially different path integrals of the neutron spin in the magnetic field and therefore the pure magnetic interaction. A similar experiment was later performed, using the high angular sensitivity of double crystal diffractometer, reconstructing the static magnetic field due to refraction by a magnetic field (Strobl et al. 2007). Modern instruments, such as PONTO I and PONTO II (Treimer et al. 2011), do not differ principally from the ‘E8’ instrument. Present-day instruments use a white beam and polarize neutrons with He<sup>3</sup> filters (Surkau et al. 1997), but a detailed spin analysis of transmitted and depolarized neutrons is quite difficult because only a 2D-position resolved depolarization of the polarized neutron beam behind the sample can be determined. Instruments at pulsed sources, such as PSI, SNS, IPARC (Shinohara et al. 2011) or ISIS, also use polarized neutrons for imaging, and in the future there will be one or two instruments for radiography and tomography with polarized neutrons at the European Spallation Source (ESS).

Polarized neutron imaging opens a large number of new fields of applications not only in physics but also in geology, archaeology, cultural heritage, applied materials research and of course, especially, magnetism. Magnetic fields, magnetic domains and quantum effects can be visualized and quantified in the bulk of matter. One of the top applications of polarized neutron imaging up to now is superconductivity, where the Meissner effect, intermediate state and flux pinning and trapped magnetic fields can be visualized and quantified.

## Polarizing Neutrons

There are different methods to polarize a neutron beam (note, polarization is a many particle parameter), to measure and to change the polarization of a beam. Commonly known methods are magnetic crystal Bragg diffraction, total reflection on super-mirrors (Krist et al. 1998; Boeni et al. 2009) and He<sup>3</sup> spin filters (Surkau et al. 1997). A special method uses Zeeman-splitting by wedge shaped magnetic (air) prism and the high angular sensitivity of a double crystal diffractometer (Badurek et al. 2000; Badurek 2011).

Magnetic crystal Bragg diffraction utilizes that the cross section ( $d\sigma/d\Omega$ ) of a magnetic crystal depends on two parameters  $b$  and  $p$

$$\left(\frac{d\sigma}{d\Omega}\right)_{\pm} \propto (b \pm p)^2, \quad (62)$$

where  $b$  ( $b_{coh}$ ) and  $p$  are the coherent nuclear and magnetic scattering lengths, respectively. They are physical constants of elements that must be measured by experiments (e.g. by measuring the index of refraction  $n$  of an element (see Eq. 64) (for non-magnetic values see, e.g. (Sears 1992), for magnetic form factors see (Magnetic form factors)).

There are a few crystals,  $\text{Fe}_3\text{Si}$ ,  $\text{Co}_{92}\text{Fe}_8$ , where in the presence of an external magnetic field  $b$  and  $p$  are of the same size ( $p \cong b$ ) (Shull et al. 1951), so that the two cross sections become

$$\begin{aligned} \left(\frac{d\sigma}{d\Omega}\right)_+ &\propto (b+p)^2 = (b+b)^2 = 4b^2 \\ \left(\frac{d\sigma}{d\Omega}\right)_- &\propto (b-p)^2 = (b-b)^2 = 0 \end{aligned} \quad (63)$$

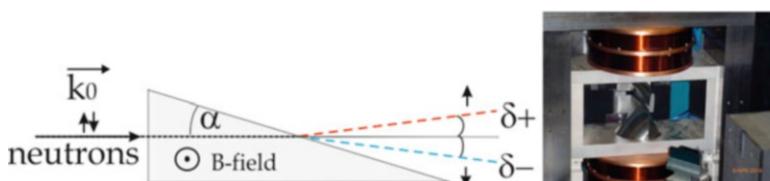
In that case, the Bragg-reflected ( $b=p$ ) neutron beam is polarized with the scattering amplitude  $4b^2$  having spin up and spin down scattered neutrons have an amplitude equal zero; so the Bragg-reflected beam is polarized (Bacon 1975) (section “[Neutron Grating Interferometry](#)” in the ‘Bacon’ is an old but very instructive introduction). Compounds like  $\text{Co}_{92}\text{Fe}_8$  ( $d_{200} = 0.1767$  nm),  $\text{Fe}_3\text{Si}$  ( $d_{111} = 0.3263$  nm) and  $\text{Cu}_2\text{MnAl}$  ( $d_{111} = 0.3435$  nm) yield the best neutron polarization.

Another method to polarize neutrons is total reflection using (Eq. 64). The index of refraction is different for spin up neutrons and spin down neutrons, and causes different angles of total reflection

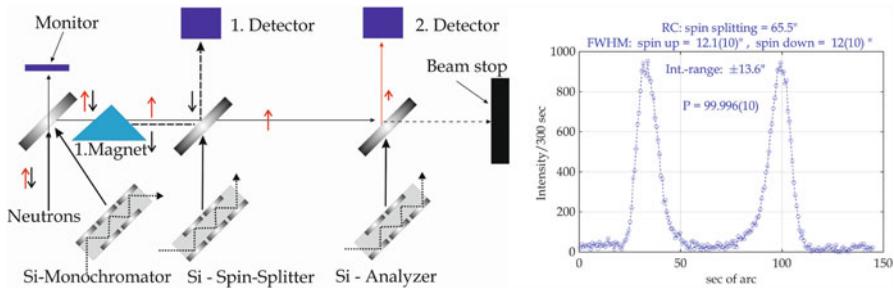
$$n = 1 - \lambda^2 \left( \frac{Nb_{\text{coh}}}{2\pi} \mp \frac{m\mu B}{h^2} \right), \quad (64)$$

Therefore, the critical angle of total reflection  $\alpha_{\text{crit}}$  is also different for spin up and spin down neutrons.  $\alpha_{\text{crit}}$  is given by  $\alpha_{\text{crit}} = a \cos(n)$ . Assuming the nuclear part in (Eq. 64) to be  $1 \cdot 10^{-5}$ ,  $B = 1$  T (magnetic field of saturation) and  $\lambda = 0.3956$  nm, the glancing angle  $\alpha_{\text{crit}} \uparrow = 0.32^\circ$  for spin up neutrons and  $\alpha_{\text{crit}} \downarrow = 0.167^\circ$  for spin down neutrons, which can be used to obtain a polarized beam for glancing angles  $\alpha > 0.167^\circ$ .

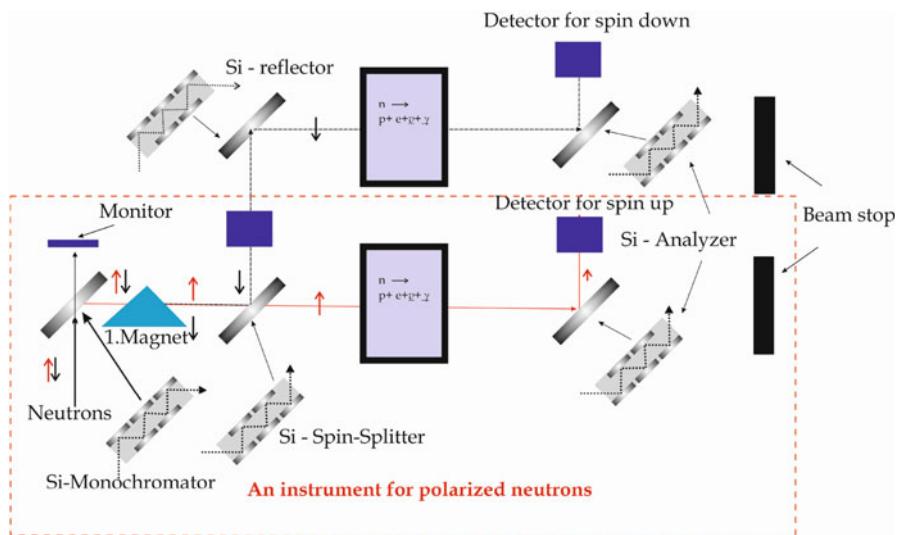
Another method uses Zeeman splitting. Figure 17 shows the geometry for spin-dependent refraction by an air magnetic wedge, and Fig. 18 shows the setup ‘Teststand’ at the BER II and measurements of the beam polarization by spin-splitting magnetic wedge. A similar layout would serve two pure spin states simultaneously as shown in Fig. 19. To date, it is the only technique to get a 100% beam polarization. The overlap of spin up and spin down neutrons is less than  $< 10^{-9}$  for a beam separation of only  $2 \times 13.5 = 27$  s of arc based on dynamical neutron scattering by perfect crystals. In Fig. 18 the peak separations was  $> 65$  s of arc and the polarization could be determined with  $P = 99.996(10)$ ; however, taking into account the overlap of both peaks with less than  $10^{-13}$ , no spin down neutron will be found in the spin up peak and vice versa.



**Fig. 17** Magnetic air wedge for spin splitting, left: ray geometry, right: magnet as used in Teststand at BER II



**Fig. 18** Left: Layout of the instrument ‘Teststand’ at BER II, right: spin-splitting by magnet, rocking curve with spin-splitter, beam polarization  $P = 99.996(10)\%$ . Note, the overlapping intensity of spin up and spin down beam for this beam separation is  $<10^{-13}$ !



**Fig. 19** Possible layout of an instrument that uses both spin up and spin down, note the possibilities this instrument provides. With the spin-splitter one can tune the spin orientation for the lower and upper beam, etc

Other methods to polarize neutrons use multilayers (supermirrors) (Mezei 1972; Schaerpf 1989; Hayter and Mook 1989; Krist et al. 1998; Boeni et al. 2009) and polarized  $^3\text{He}$  spin filters. Supermirrors also work with total reflection, but on appr. 1000 layers that differ in thickness  $d_{1\dots 1000}$  and yield an enlarged critical angle  $m \cdot \alpha_{\text{crit}}$ ,  $m = 3 - 4$ . (Hayter and Mook 1989; Krist et al. 1998; Boeni et al. 2009).

Recently, nuclear spin polarized  $^3\text{He}$  produced by optical pumping has been used as a spin filter and become important especially in neutron imaging, because different from the methods described-above, spin polarized  $^3\text{He}$  can polarize a white beam, thus gaining two or three orders of magnitude intensity (Surkau et al. 1997). One uses the special reaction



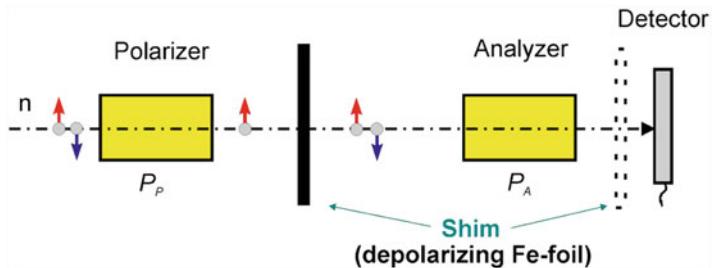
The absorption cross section for neutrons with spins anti-parallel to those of polarized  ${}^3\text{He}$  nuclei is about three orders of magnitude larger than if neutron and  ${}^3\text{He}$  spins are parallel. Thus, the transmitted neutron beam contains nearly spin polarized neutrons. However, the polarization is much less than 90%, and there is certain decay of polarization in time that has to be taken into account for long-time experiments.

## Determination of Beam Polarization

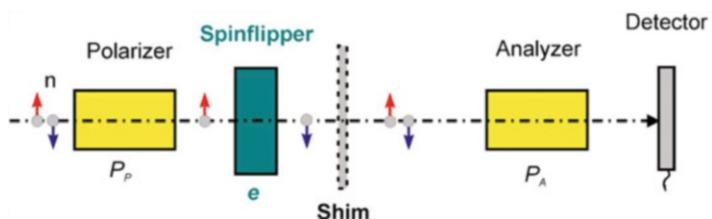
Working with polarized neutrons involves a description of methods that determine the polarization  $P$  of the neutron beam. The problem here is the lack of a universal probe (standard) that measures a beam polarization by 100%. Important developments in ‘working with spin’ (spin rotation) were made by (Mezei 1972, 1976) and (Schaerpf 1989; Hayter and Mook 1989). Thus, the ‘shim’-method and the ‘spin flipper’ method became established parts of experiments with polarized neutrons. However, there are other, sophisticated methods, such as the ‘double flipper method’ and the ‘combined static and pulse flipper technique’ (Badurek 2011) (Figs. 20 and 21).

The polarization of a neutron beam is a many-particle feature, defined as

$$P = \frac{N_{\uparrow} - N_{\downarrow}}{N_{\uparrow} + N_{\downarrow}} \quad (66)$$



**Fig. 20** Shim method, after G. Badurek (2011), see text



**Fig. 21** Single flipper method, after G. Badurek (2011), see text

where  $N_\uparrow$  and  $N_\downarrow$  are the number of spin up neutrons and spin down neutrons, respectively. The polarization  $P$  of the neutron beam can be derived from the flipping ratio  $F = N_\uparrow/N_\downarrow$  as

$$\frac{N_\uparrow - N_\downarrow}{N_\uparrow + N_\downarrow} = \frac{N_\uparrow/N_\downarrow - 1}{N_\uparrow/N_\downarrow + 1} = \frac{F - 1}{F + 1} \Rightarrow P = \frac{F - 1}{F + 1} \quad (67)$$

The ‘shim’ method first places a depolarizing Fe–foil between polarizer and analyser ( $I_1$ ), and then between analyser and detector ( $I_2$ ), both background corrected. Then

$$P = \sqrt{\frac{I_1 - I_2}{I_2}} = \sqrt{S - 1} \quad (68)$$

the ratio  $S = I_1/I_2$  is called ‘shim ratio’.

A second method to determine the polarization of a neutron beam uses a spin-flipper ( $\pi$ -flipper) placed between polarizer and analyser (‘single flipper method’). There the flip-ratio  $R = \frac{I}{I_F}$  is written as (Badurek 2011)

$$R = \frac{I}{I_F} = \frac{1 + P_P \cdot P_A}{1 - e \cdot P_P \cdot P_A} \cong \frac{1 + P^2}{1 - e \cdot P^2} \quad (69)$$

$$e \equiv 2k - 1 = \frac{R - S}{R \cdot (S - 1)}$$

with  $I$  = flipper off and  $I_F$  = flipper on. The flip-efficiency ‘ $e$ ’ is defined as  $2k - 1$ , with ‘ $k$ ’ as the spin-flip probability.

## Interaction with Magnetic Fields

The description of a neutron spin in a magnetic field can be treated rather classically (see Mezei 1972, 1976; Schaerp 1989). In Quantum mechanics, the spin  $\mathbf{S}$  is connected with the angular momentum  $\mathbf{l}$  as

$$\vec{\mu}_l = g_l \mu_k \cdot \frac{\vec{l}}{\hbar} \quad \text{and} \quad \vec{\mu}_s = g_s \mu_k \cdot \frac{\vec{S}}{\hbar} \quad (70)$$

where  $\mu_k$  is

$$\mu_k = \frac{e \cdot \hbar}{2 \cdot m_p} = 3.1525 \cdot 10^{-14} \frac{MeV}{T} \quad (71)$$

$m_p$  is the mass of a proton and  $\mu_k$  is called the nuclear magneton ( $\mu_k = 5.05078343 \times 10^{-27} [\text{J/T}]$ );  $\hbar = 6.6260755 \times 10^{-34} [\text{J.s}]$  is the Planck constant,  $\hbar = h/2\pi$ . The  $g_{l,s}$  are the g-factors of the proton and neutron. For protons  $g_{p,l} = 1$  and  $g_{p,s} = 5.5857$  and for neutrons  $g_{n,l} = 0$  and  $g_{n,s} = -3.826085$ . Thus,

the magnetic moment of the neutron  $\mu_n = -1.9130427\mu_k$ , i.e.  $\mu_n = -9.662364 \times 10^{-27}[\text{J/T}]$ . The time-dependent motion of the spin of a neutron  $\mathbf{s} = \mathbf{s}(t)$  in a magnetic field  $B$  is described by the equation

$$\frac{d\mathbf{S}(t)}{dt} \Big|_j = \frac{\mu_k}{\hbar} g_n [\mathbf{S}(t) \times \mathbf{B}(t)]_j \quad j = x, y, z \quad (72)$$

Consider a rectangular coordinate system  $\{x, y, z\}$ , where the magnetic field direction of  $\mathbf{B}$  is parallel to the  $z$ -axis (see small sketch Fig. 22). If the neutron spin lies in the  $\{x, y\}$  plane, its motion around  $\mathbf{B}$  is given by

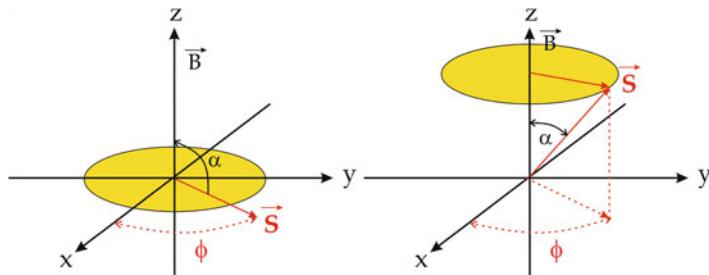
$$\mathbf{S}(t) = S \begin{pmatrix} \cos(\phi(t)) \\ \sin(\phi(t)) \\ 0 \end{pmatrix} \quad (73)$$

The rotation of  $\mathbf{S}$  around  $\mathbf{B}$  is shown in Fig. 22. Independent of the angle  $\alpha$  between  $\mathbf{S}$  and  $\mathbf{B}$  the spin rotates with the so-called Larmor frequency  $\omega_L = \gamma_L \cdot B$ , where  $\gamma_L$  is the gyro-magnetic ratio of the neutron, given as  $\gamma_L = g_{n,s} \cdot \mu_k/\hbar = -1.8324698985 \times 10^8 \text{ rad} \cdot \text{s}^{-1} \cdot \text{T}^{-1}$ . From the Larmor frequency  $\omega_L$  one can calculate the rotation angle  $\phi$  as

$$\phi = \omega_L \cdot t = \gamma_L \cdot B \cdot t, \quad (74)$$

$t$  is the transit time of the neutron through the magnetic field  $\mathbf{B}$ , actually only the amount of  $\mathbf{B} = B$  is important, not the orientation of  $S$  with respect to  $B$ . With the known velocity (wavelength) of the neutron, given by the de Broglie relation  $\lambda = h/p$ ,  $p = m_n \cdot v$ , the neutron velocity  $v$  becomes

$$v = \frac{3.956034 \cdot 10^{-7} [\text{m}^2 \cdot \text{s}^{-1}]}{\lambda [\text{m}]} \quad (75)$$



**Fig. 22** Spin rotation  $\mathbf{S}$  in a constant magnetic field  $\mathbf{B}$ . Independent of the angle  $\alpha$ ,  $\mathbf{S}$  rotates around  $\mathbf{B}$  with  $\omega_L$

For example, the velocity  $v$  of neutrons with the wavelength of  $\lambda = 0.1$  nm is 3956 m/s. The rotation angle  $\phi$  in a field  $B$  is determined by the path length ‘ $s$ ’ of the neutron in  $B$ , i.e. by the path integral

$$\phi = \frac{\gamma_L}{v} \int B \cdot ds = \frac{\gamma_L \cdot m}{h} B \cdot s \cdot \lambda \quad (76)$$

The factor  $\frac{\gamma_L \cdot m}{h}$  is constant  $= 4.632088 \cdot 10^{14} [\text{T}^{-1} \text{m}^{-2}]$ , therefore for, e.g.  $B = l$  mT and  $s = 0.001$  m, and  $\lambda = 0.3956034$  nm,  $\phi = 4.6320876 \cdot 10^{14} \text{ T}^{-1} \cdot \text{m}^{-2} \times 3.956034 \cdot 10^{-10} \text{ m} \cdot \text{s}^{-1} = 0.183247$ [rad], which corresponds to angle  $\phi = 10.5^\circ$ . A spin-flip ( $\phi = 180^\circ$ ) would occur for a path length  $\sim 17.15$  mm. With these numbers one can check the sensitivity detecting magnetic fields in matter. One must keep in mind, that (usually) magnetic fields have no strict boundaries and the interaction with a neutron spin must consider paths where  $\omega_L \geq \omega_B$  rotation frequency of  $\mathbf{B}$  and paths where  $\mathbf{B}$  abruptly changes its orientation. If  $\omega_L \geq \omega_B$  the neutron spin will follow (adiabatically)  $\mathbf{B}$ , for any other case  $\mathbf{S}$  rotates around  $\mathbf{B}$ .

## Calculation of Magnetic Fields

With new instruments and techniques, information obtained from radiographies and tomographies increased from simple observations of matter to very sophisticated and challenging physics. Research work focused on investigation of magnetic fields (Manke et al. 2009), magnetic domains (Manke et al. 2010; Schulz et al. 2010a, b), quantum phases and quantum criticality (Schulz 2010) or trapped fields in superconductors (Treimer et al. 2012a, b). Polarized neutron imaging (PNI) requires the selection of one spin state in front of the sample, and a spin analysis in order to determine the change of this spin state by the sample. This can be done by different techniques, depending on the use of monochromatic neutrons or of the white spectrum. The spatial resolution in PNI is – to date – of the order of 100  $\mu\text{m}$  or even less; thus, the field of applications will cover a larger number of topics than known today, ranging from industrial demands to problems of fundamental physics.

The detection of magnetic fields in matter uses the interaction of the neutron spin  $\mathbf{S}$  with the magnetic field  $\mathbf{B}$ . In some cases (e.g. trapped fields in superconducting samples) the strength of  $\mathbf{B} = B$  in matter is much larger than the surrounding stray field and one can assume, that the interaction of  $S$  with the stray field  $B$  is an adiabatic process, if  $\mathbf{S}$  is parallel (or nearly parallel) to  $\mathbf{B}$  or the change of  $\mathbf{S}$  due to a small  $|\mathbf{B}|$ .  $\mathbf{S}$  follows  $\mathbf{B}$ , if  $\omega_L \gg \omega_B$  ( $\omega_B$  = rotation frequency of  $\mathbf{B}$ ), i.e. the change of the orientation of  $\mathbf{B}$  is small. If  $\mathbf{S}$  is not parallel to  $\mathbf{B}$ , it performs Larmor precessions, as given in (Eq. 74).

In materials science, the study of domain structures in ferromagnetic samples of bulk materials by neutron scattering is of great interest. Their study, based on the neutron depolarization concept in magnetic structures, was first introduced by Halpern and Holstein (Halpern and Johnson 1939). They proposed that the polarization of a neutron beam rotates on its way through a sample with a given magnetic

field, and with their simple experiment they were able to yield information about the mean domain size in the sample. In 1973, Rekveldt expanded this concept for three dimensional techniques (Rekveldt 1973) and was able to give information about the correlation of the domain magnetization. This work was then propagated in 1996 (Hochhold et al. 1996; Badurek et al. 1997a). The concept is similar to the common X-ray technique but far more complex when it comes to the evaluation of the data. This is because of the tensorial description of the neutron spin.

Other methods to determine magnetic fields have the mathematical tomographic background of ART, SIRT and SART, which are implemented, e.g. in MATLAB, and so one is able to calculate the behaviour of the polarization of a neutron beam in order to evaluate depolarization measurements. Varying the incident direction of the neutron beam and registering as many projections as are necessary, one can perform proper reconstructions. For each beam one can get a depolarization matrix, which contains information on the magnetic structures along the beam. The beams are considered very thin and the effects of the Bloch walls at the domain borders are neglected. With the depolarization matrix for each beam, one can get tomographic information. Because of the matrix structure and not just scalar quantities directly related to the line integrals, it is possible to reconstruct every voxel in the sample in three dimensions.

This method by simulation of different kinds of domain structures in ferromagnetic samples is given below. With these depolarization data, one can compare different approaches on iterative algorithms to reconstruct the structures of two and three dimensional objects.

## Neutron Polarization

(This part was mathematically worked out by Johannes Nicol)

The depolarization of a neutron beam by transmission through a magnetic sample can be described by the semiclassical spin rotation formalism (see e.g. (Schaerpf)).

The polarization of a neutron beam can be derived from single neutron spin rotation as shown in Fig. 11 or one can consider it (as it actually is) as a many particle parameter  $\mathbf{P}$ . The equation of motion is equivalent to (Eq. 72)

$$\frac{d\mathbf{P}}{dt} = \gamma_L \mathbf{P} \times \mathbf{B}, \quad (77)$$

where  $\gamma_L$  is the gyromagnetic factor of the neutrons, but (Eq. 77) is only valid if the effect of, e.g. the domain walls on  $P$  can be neglected. The time dependence of  $\mathbf{P}$  in (Eq. 77) can be written as  $z$ -coordinate dependence due to  $z = z_0 + v \cdot t$ . Neutrons moving in the  $z$ -direction with velocity  $v$  can now be written as matrix equation (Rekveldt 1973; Leeb et al. 1998; Badurek et al. 1998),

$$\frac{d\mathbf{P}(z)}{dz} = \mathcal{A}(z) \cdot \mathbf{P}(z). \quad (78)$$

where the cross product in (Eq. 77) is replaced by  $\mathcal{A}(z) \cdot \mathbf{P}(z) \cdot \mathcal{A}(z)$  contains the velocity  $v$  of the neutrons and the structure of the magnetic probe along the neutron path. The items of  $\mathcal{A}(z)$  are the components of the magnetic flux  $B = (B_x, B_y, B_z)$  as follows (Rekveldt 1973)

$$\mathcal{A} = \frac{\gamma_L}{v} \begin{pmatrix} 0 & B_z & -B_y \\ -B_z & 0 & B_x \\ B_y & -B_x & 0 \end{pmatrix}. \quad (79)$$

That is, the change from  $\mathbf{P}_o$  to  $\mathbf{P}(z)$  can be described applying a rotation matrix to  $\mathbf{P}_o$ . If one considers a homogeneous magnetic field  $\mathbf{B}$  then  $\mathbf{P}$  rotates around  $\mathbf{B}$  through an angle  $\phi = \gamma_L \cdot B \cdot t = \gamma_L/v \cdot (z - z_0) \cdot B$  (cp. (Eq. 74)) and (Eq. 78) can be rewritten as (Leeb et al. 1998)

$$\mathbf{P}(z) = \mathcal{D}(\mathbf{n}, \phi) \mathbf{P}(z_0) \quad (80)$$

where  $\mathcal{D}$  is also a rotation matrix and  $\mathbf{n}$  is the axis of rotation  $\mathbf{n} = \mathbf{B}/|\mathbf{B}|$  (see Fig. 23),  $\mathbf{P}(z)$  becomes

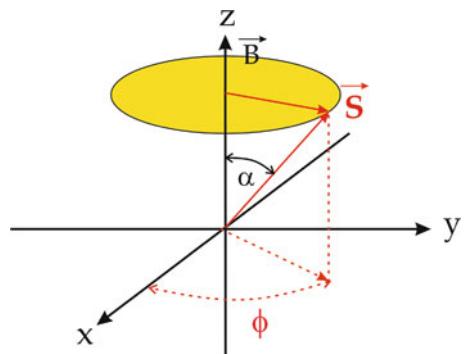
$$\mathbf{P}(z) = \exp\left(\int_{z_0}^{\rightarrow z_e} dz \cdot \mathcal{A}(z)\right) \cdot \mathbf{P}_o. \quad (81)$$

$z_0$  and  $z_e$  characterize the entrance and the exit points of the beam. The path ordering is indicated by the arrow on the integral. Note that  $\mathcal{A}(z)$  has non-Abelian (non-commutative) properties (Leeb et al. 1998). With the dot product of the initial  $\mathbf{P}_o$  and final  $\mathbf{P}(z)$  polarization, one can calculate the cosine of the angle  $\phi$  between them and so the resulting change of the polarization.

$$\mathcal{D} = \exp\left(\int_{z_0}^{\rightarrow z_e} dz \cdot \mathcal{A}(z)\right) \quad (82)$$

Considering infinitely thin beams, the matrix  $\mathcal{D}$  is a pure rotation matrix, which does not change the length of the polarization vector. The approach for the tomographic inverse problem is the solution of (Eq. 82) with the important feature of path ordering.

**Fig. 23** Rotation of  $\mathbf{S}$  around  $\mathbf{B}$  in an  $\{x, y, z\}$  coordinate system



Following (Eq. 82),  $\mathcal{D}$  must be factorized to get a solution.  $\mathcal{D}$  can be written as a product

$$\mathcal{D} = \mathcal{L} \cdot \mathcal{K}, \quad (83)$$

where  $\mathcal{L}$  contains the line integrals of every projection (and thus magnetic structures along the path) and  $\mathcal{K}$  is defined as a correction matrix. Therefore,  $\mathcal{L} = \exp\left(\int_{z_0}^{z_e} dz \cdot \mathcal{A}(z)\right)$  yields the magnetic distribution when Radon transformation is applied to line integrals. For this (Badurek et al. 1998) presents an iterative procedure for calculating a given magnetization distribution for each path value  $\mathcal{D}_{\text{theo}}$  and  $\mathcal{L}_{\text{theo}}$  and  $\mathcal{K}_{\text{theo}} = \mathcal{L}_{\text{theo}}^{-1} \mathcal{D}_{\text{theo}}$ . With  $\mathcal{D}$  one gets  $\mathcal{K}_{\text{exp}}$  from the line integrals  $\mathcal{K}_{\text{theo}} = \mathcal{L}_{\text{theo}}^{-1} \mathcal{D}_{\text{theo}}$ .

Applying the Radon transform to the line integrals  $\mathcal{L}_{\text{exp}}$  the assumed magnetization distribution appears improved and can be used as input for a next approach.

As mentioned above,  $\mathcal{L}$  contains the line integrals of every projection (and thus magnetic structures along the path). If one neglects  $\mathcal{K}$ , and considers only the information of  $\mathcal{L}$  for each projection the tomographic problem simplifies to three independent scalar integrals,

$$\int_{z_0}^{z_e} dz B_k^{(l)} = \frac{\nu}{\gamma_L} (\ln \mathcal{D}_{i,j}), \quad \begin{cases} k = x & i, j = y, z \\ k = y & i, j = z, x \\ k = z & i, j = x, y \end{cases} \quad (84)$$

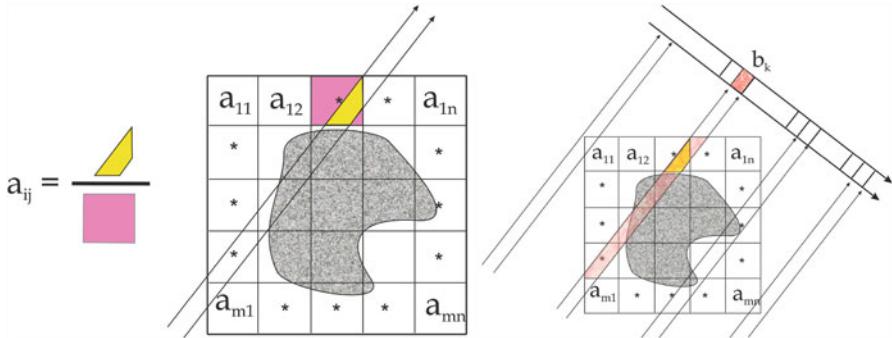
This can be inverted individually by Radon Back-Transformation techniques as well as iterative algorithms (Leeb et al. 1998; Badurek et al. 1998). In the next chapter three basic concepts of iterative methods will be described, the algorithms of Kaczmarz, SIRT and SART.

## Reconstruction Methods

### Art

One of the oldest reconstruction techniques is based on linear algebra. Consider a  $m \times n$  matrix and allocate each item an unknown  $x_J$ ,  $J = 1 \dots N = m \times n$ . The scanning is shown in Fig. 24,  $b_k$  is the sum of all contributing weighted pixel of the  $k$ -th ray ( $k = 1 \dots K$ ). Scanning the pixel field ‘tomographically’, (see section “Mathematics of Computerized Tomography,” Fig. 7), i.e. the pixel field is scanned under angles  $\theta$ ,  $0 \leq \theta \leq \pi$ , one gets ‘projections’ (similar  $P(\theta, t)$ ), and so a system of linear equations as

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_N &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_N &= b_2 \\ \vdots & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_N &= b_m \end{aligned} \quad (85)$$



**Fig. 24** Algebraic Reconstruction Technique (ART), left: ray traverses a pixel field and hits different pixel  $a_{ij}$ , the amount of contribution of  $a_{ij}$ , yielding  $b_k$

(Equation 85) can be expressed as matrix equation,

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b}, \quad (86)$$

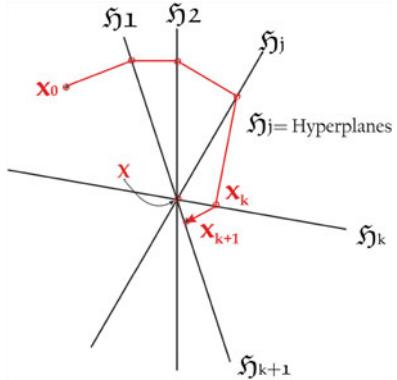
where  $\mathbf{A}$  is a full rank  $m \times n$  matrix with  $m \geq n$  and  $\mathbf{b} \in \mathbb{C}^m$  and  $\mathbf{x}$  contains the  $N$  unknowns of the pixel field. In order to get an idea of (Eq. 85), let the pixel field – for simplicity – have  $100 \times 100$  items; thus,  $N = \text{number of unknown} = 10$ . (Brenizer 2013). For  $10^4$  unknowns one should have  $10^4$  linear independent equations, which involves the calculation of  $10^8 a_{ij}$ . In the past it was a problem to solve such a large system of equations because of limited computing power, and therefore the underlying pixel field was limited to rather small pixel arrays (Treimer et al. 1991).

The most common method to solve such large equation systems uses algebraic reconstruction techniques; the primal is called the Kaczmarz algorithm (Kaczmarz 1937). It is an iterative algorithm and one of the most popular solvers for under-determined systems of linear equations (“ill-posed problems”). The idea is shown in Fig. 25. From an arbitrary starting point  $x_0$ , the nearest point on a hyperplane  $\mathcal{H}_1$  is the next starting point for a projection on  $\mathcal{H}_2$ . Each hyperplane  $\mathcal{H}_i$  is an equation in (Eq. 85). In the ideal case, i.e. in the case of  $N$  independent linear equations, the solution is the intersection of all  $\mathcal{H}_i$  (The intersection of two straight lines is the solution of their corresponding two equations with two unknowns. Note that both straight lines must be in the same plane). That means one uses one equation in each step and sweeps through the rows of  $\mathbf{A}$  in a cyclic manner, projecting the last iterate orthogonally onto the solution hyperplane of  $\langle a_i, x \rangle = b_i$ .

The solution is taken as the next iterate and updates  $x^k$  in the following way:

$$\begin{aligned} x_0^k &= x^k \\ x_i^{k+1} &= x_{i-1}^k + \lambda \sum_{i=1}^n \frac{b_i - \langle a_i, x_{i-1}^k \rangle}{\|a_i\|_2^2} a_i \quad i = 1, 2, \dots, m \\ x_m^{k+1} &= x_m^k \end{aligned} \quad (87)$$

**Fig. 25** Lines represent hyperplanes in (Eq. 85)



Actually, the number of linear independent equations in (Eq. 85) can be much less than  $N$ . It could be shown that even with an extreme ill-posed problem (half the number of linear independent equations) quite good reconstruction could be realized (Treimer et al. 1991; Maass et al. 1992; Treimer and Feye-Treimer 1998). However, if the linear system is consistent, Kaczmarz's method converges to a least squares solution of this system (Treimer et al. 2005b). The solution is found by comparing the last solution with the previous one. A numerical limit  $L < \|x^i - x^{i-1}\|$  determines the accuracy of the reconstruction. One must keep in mind that the found solution is a rather small ( $N - \Delta$ ,  $\Delta \geq 0$ )-dimensional volume and  $\mathbf{x}$  is found by a least squares solution. The solution vector  $\mathbf{x}$  is then mapped on a two-dimensional matrix, yielding the two-dimensional reconstructed image.

A solution can also be found by means of Lagrangian multipliers. Without going into the details, the solution vector  $x$  can be found by

$$\mathbf{x} = A^T \cdot [A \cdot A^T]^{-1} \cdot \mathbf{b} \quad (88)$$

where  $A^T$  is the transposed matrix of  $A$ .

Simple example:

Given a pixel field as a  $3 \times 3$  matrix with nine unknowns  $x_1 \dots x_9$ . The field was scanned by five projections, yielding values given in  $\mathbf{b}$ , the  $a_{ij}$  are the weighted contributions of a pixel as shown in Fig. 24.

$$A = \begin{pmatrix} 0.971 & 0.702 & 0.104 & 0.466 & 0.049 & 0.117 & 0.740 & 0.0044 & 0.956 \\ 0.751 & 0.678 & 0.663 & 0.071 & 0.861 & 0.889 & 0.152 & 0.574 & 0.175 \\ 0.834 & 0.859 & 0.983 & 0.709 & 0.816 & 0.974 & 0.314 & 0.206 & 0.948 \\ 0.25 & 0.866 & 0.206 & 0.753 & 0.375 & 0.631 & 0.736 & 0.389 & 0.98 \\ 0.328 & 0.627 & 0.490 & 0.268 & 0.065 & 0.454 & 0.647 & 0.266 & 0.625 \end{pmatrix}, \quad b = \begin{bmatrix} 5.114 \\ 24.858 \\ 35.205 \\ 18.6 \\ 48.05 \end{bmatrix},$$

using (Eq. 88) yields  $\mathbf{x} = \begin{bmatrix} -17.586 \\ 7.348 \\ 61.316 \\ -23.427 \\ -33.288 \\ 16.643 \\ 27.48 \\ 6.299 \\ 0.975 \end{bmatrix}$ . Checking the result by calculating  $\text{vec}A \cdot \mathbf{x}$  one gets

$$A \cdot \mathbf{x} = \begin{bmatrix} 5.114 \\ 24.858 \\ 35.205 \\ 18.6 \\ 48.05 \end{bmatrix} = b \text{ (see above), } \mathbf{x} \text{ becomes the reconstructed image matrix if}$$

it is written as

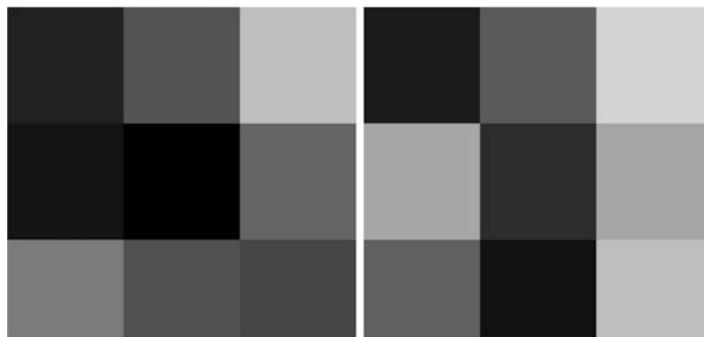
$$\mathbf{x} \rightarrow \mathbf{B}_0 = \begin{bmatrix} -17.586 & 7.348 & 61.316 \\ -23.427 & -33.288 & 16.643 \\ 27.48 & 6.299 & 0.975 \end{bmatrix}.$$

Because all items must be positive (there are no negative intensities) one must add a ‘background matrix  $\mathbf{B}_{\text{bg}}$  to  $\mathbf{B}_0$ ,  $\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_{\text{bg}}$ , and scale  $\mathbf{B}$  by multiplying with a factor (e.g. 2 or larger) in order to enhance the image contrast.

$$\mathbf{B}_{\text{Gb}} = \begin{bmatrix} 111 \\ 111 \\ 111 \end{bmatrix} \cdot 34 \quad \Rightarrow \mathbf{B} = \begin{bmatrix} 33 & 83 & 191 \\ 21 & 1 & 101 \\ 123 & 81 & 70 \end{bmatrix}.$$

$\mathbf{B}$  becomes an image with scaling factors 2 and 64, see Fig. 26.

Note, adding and scaling may warp the result, which may lead to the wrong interpretation of the image.



**Fig. 26** Left image:  $\mathbf{B}$  scaled by a factor = 2, right image: scaling factor = 64. By scaling one can stretch the gray values in order to achieve an improved appearance

## SIRT and SART

This part was mainly worked out by Johannes Nicol.

Unlike ART, ‘Algebraic Reconstruction Technique’, ‘Simultaneous Iterative Reconstruction Technique’ (SIRT) uses all  $a_{i,k}$  simultaneously (Andersen and Kak 1984; Van Hemelryk 2007; Maria and Hansen 2010; Hansen and Saxild-Hansen 2012) and (Eq. 87) is modified to (Eq. 89).

$$x^{k+1} = x^k + \lambda_k A^T (b - Ax^k) \quad k = 0, 1, \dots, \quad (89)$$

where  $\lambda_k$  are relaxation parameters. The convergence is very low and it needs many more iterations (Maria and Hansen 2010), but the idea makes use of relaxation parameters, which are implemented in the Simultaneous Algebraic Reconstruction Technique (SART). SART was introduced in 1984 and it combines all advantages of ART and SIRT (Andersen and Kak 1984). In general it can be written in SIRT form, and therefore one can be categorized as a SIRT method. The SART algorithm can be written in the following form (cp. (Eq. 89), (Eq. 87) see (Hansen and Saxild-Hansen 2010; Hansen and Saxild-Hansen 2012)):

$$x^{k+1} = x^k + \lambda_k V^{-1} A^T W^{-1} (b - Ax^k) \quad k = 0, 1, \dots \quad (90)$$

Now  $A^T$  is replaced by  $V^{-1} A^T W^{-1}$ , where  $V$  and  $W$  are diagonal matrices defined as the row and column sums (Hansen and Saxild-Hansen 2012),

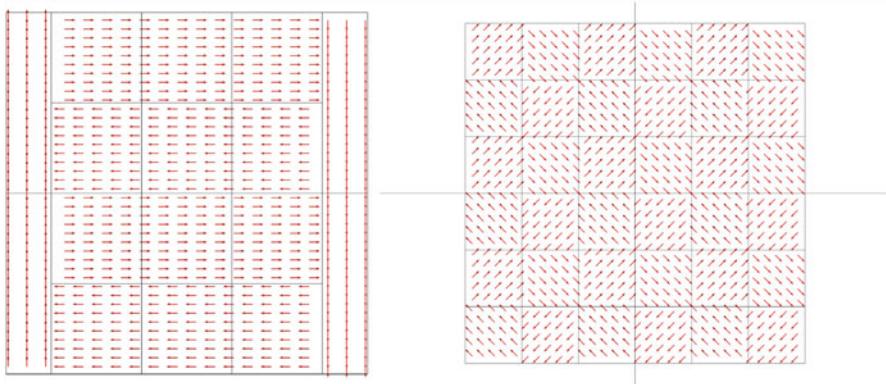
$$\begin{aligned} V &= \text{diag}\left(\|a^i\|_j\right) = \text{diag}(\varsigma^i); & \varsigma^i &= \sum_{j=1}^n a_j^i & i &= 1, \dots, m \\ W &= \text{diag}\left(\|a_j\|_1\right) = \text{diag}(\varsigma_j); & \varsigma_j &= \sum_{i=1}^m a_i^j & j &= 1, \dots, n \end{aligned} \quad (91)$$

In the following, ART, SIRT and SART are applied to a magnetic system (magnetic domains).

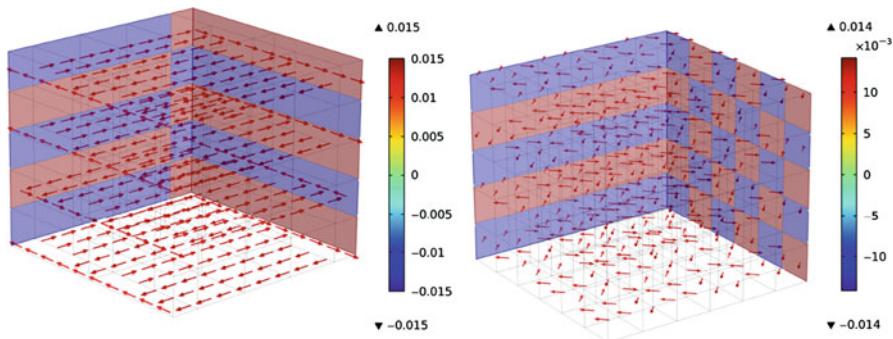
### Example: Magnetic Domains

In this example the iterative algorithms are applied to systems of magnetic domains, i.e. a set of simulated known field distribution. The cross-section of the simulated magnetized probe is divided into  $24 \times 24$  quadratic pixels in one plane. The pixel size is  $50 \mu\text{m} \times 50 \mu\text{m}$  and each pixel contains a homogeneous magnetic field component. The magnetic field distribution is divided into two different magnetic domains aligned in the x- and y-direction, respectively, while  $B_z = 0$ . Therefore, one can simulate two different kinds of magnetic samples, one sample with  $180^\circ$  magnetic domains and one with  $90^\circ$  magnetic domains. Each sample has the size of  $(1.2 \text{ mm} \times 1.2 \text{ mm} \times 1 \text{ mm})$ . In the z-direction, five slices of different domains were constructed, each one with an extension of  $(1.2 \text{ mm} \times 1.2 \text{ mm} \times 0.2 \text{ mm})$  (Figs. 27 and 28).

The normalized values (Note one has a  $B_x$  and  $B_y$  in the case of  $90^\circ$  domains) in domains were fixed to  $B_s = 15 \text{ mT}$  for the simulated  $180^\circ$  domains and to  $B_s = 20 \text{ mT}$



**Fig. 27** Slices of the magnetic field distribution: left with a 180° orientation and right with 90° orientation

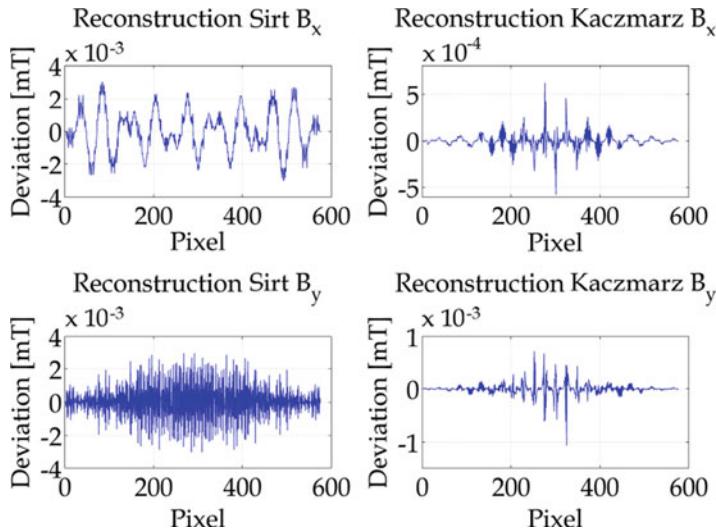


**Fig. 28** Simulated blocks: left with 180° domain orientation, right with 90° domain orientation. Reconstructed  $B_x$  and  $B_y$  values of the 180° domains and 90° domains agree with model values ( $B_s = 15$  mT of 180° domains and  $B_s = 20$  mT less than  $5 \times 10^{-3}$  mT)

for 90° domains. With such a model field distribution, one can collect a tomographic data set and apply different iterative algorithms. With 45 projections and without any assumptions about the starting domain configuration, one can achieve excellent numerical accuracy. The quality of SIRT and the Kaczmarz reconstruction technique can be controlled by plotting simple mean deviation for all 576 pixels in the sample (Fig. 29).

The deviation of accuracy for the SIRT and SART algorithms were nearly the same. The Kaczmarz method seems to yield better reconstruction for 90° and 180° magnetic domains than SIRT, for this simple model.

The size of the sample was about 1200  $\mu\text{m}$  in the y-direction and 1000  $\mu\text{m}$  in the x-direction. One can expand both techniques assuming larger pixel sizes and/or higher absolute values of  $B$  in each domain. For these cases, better techniques for finding starting values for the first assumption are needed.



**Fig. 29** Deviation of reconstruction from original  $B$ , left with SIRT, right with Kaczmarz algorithm

This is due to the fact of a failure of convergence for rotation angles larger than  $\pi$  in every direction, where one gets multiple solutions for the path ordered line integrals, which leads to a collapse of the convergence rate. Therefore, one has to find suitable ferromagnetic samples that are in a certain range of domain sizes.

Another important thing to consider is the acquired beam time for a tomography. Owing to the fact that one has to measure every projection for nine different starting vector and analyser relations, it is not suitable to use the Radon back-transformation as the reconstruction technique. It has proven more efficient to use iterative algorithms were one can solve under-determined systems of equations with a quite good accuracy. Recently by Time-of-Flight Three Dimensional Polarimetric Neutron Tomography three dimensional magnetic field strengths and directions of a test coil were reconstructed (Sales et al. 2018).

## Reconstruction of Continuous Magnetic Fields

As mentioned above, polarized neutron imaging is a perfect tool for investigations of magnetic fields in the bulk of matter. There is no other method more capable to examine 3D-position resolved magnetism inside matter. One ‘hot topic’ is the investigation of trapped fields and intermediate state in superconductors (SC). Most materials become superconducting if they are cooled down below their critical temperature  $T_c$ , i.e. they show no electrical resistance. Moreover, below external – temperature dependent – magnetic fields  $B_{\text{ext}}$  are excluded. In the case of so-called field cooling (FC), i.e. cooling in the presence of an external magnetic field  $B_{\text{ext}}$  below its critical temperature  $T_c$ ,  $B_{\text{ext}}$  is expelled (known as the Meißner–Ochsenfeld effect). However, there is no total exclusion or expulsion of the magnetic field, but

one finds a trapped (pinned) magnetic field and the occurrence of an intermediate state. This state is found in superconductor type I, where regions of normal and superconducting material coexist; however, there are hints that in the so-called Meißner phase this also can occur in superconductor type II. Trapped or pinned fields in superconductors are of special interest to understand the physics behind them and because they reduce the efficiency of all devices that use superconducting components. In this chapter, some basics of the study of trapped magnetic fields in SC I and II by means of polarized neutron imaging are presented. The main features of superconductor type I (SC I) are:

- Zero electrical resistance
- Magnetic field exclusion (temperature &  $B_{\text{ext}}$ -dependent)
- Magnetic field expulsion (Meißner-Ochsenfeld effect)
- Magnetic field trapping/pinning
- Existence of intermediate state

and of superconductor type II (SC II):

- Two superconducting states given by  $B_{C1}$  and  $B_{C2}$
- $0 < B_{C1} =$  Meißner phase,  $B_{C1} < B < B_{C2} =$  Shubnikov phase ( $B_{C2} > 100$  Tesla!)
- Shubnikov phase: magnetic fields can penetrate SC  $\leftrightarrow$  Abrikosov vortices Magnetic field trapping/pinning

Thus, there are many questions still to be answered, such as amount and shape and the thermodynamics of trapped fields. In order to understand how magnetic fields can be studied by polarized neutron imaging we refer to section “[Determination of Beam Polarization](#).” For simplicity, one considers first the B-field of a coil, as shown in Fig. 30.

For small magnetic fields (e.g. some mT), inside the coil the B-field outside of the coil can be neglected regarding a change of the neutron spin orientation. A good approximation to estimate this fact is to consider the density of flux lines inside and outside of the coil, which are closed loops. The strength of a magnetic field  $B$  inside a coil is given by  $B = \mu_0 \cdot I \cdot \frac{W}{l}$ , with  $\mu_0 = 4\pi \times 10^{-7} \frac{\text{Vs}}{\text{Am}}$ ,  $I$  = current [Ampere],  $W$  = number of windings and  $l$  = length [Meter] of the coil.

## Polarized Neutron Imaging: Radon Transform for Image Reconstruction

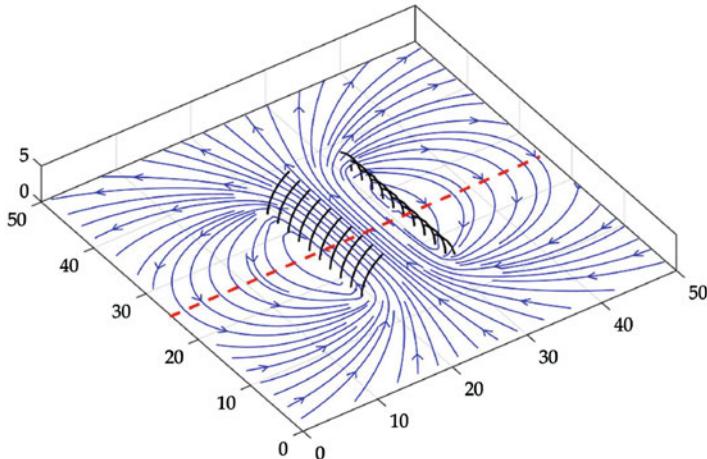
A standard set-up for polarized neutron imaging consists of a spin polarizer, sample and a spin analyser (see Fig. 31). The polarizer device in front of the sample selects neutrons with, e.g.  $|\uparrow\rangle$  (spin ‘up’) and the analyser is 100% transparent for ‘spin up neutrons’; thus, a 2D-detector behind the analyser detects all transmitted neutrons having ‘spin up’. Neutrons with spin  $\mathbf{S}$  parallel to  $\mathbf{B}_{\text{Analyser}}$  can pass the spin analyser

and reach the detector, neutrons with spin  $\mathbf{S} \nparallel \mathbf{B}$  are blocked depending on their angle  $\alpha$  with respect to  $\mathbf{B}$  (The transmitted intensity behaves similar to polarized light passing two horizontally oriented polarizers, which block light if they are perpendicular to each other, different from neutrons (Fermions), where the polarizer must be orientated opposite the analyser to block neutrons.).

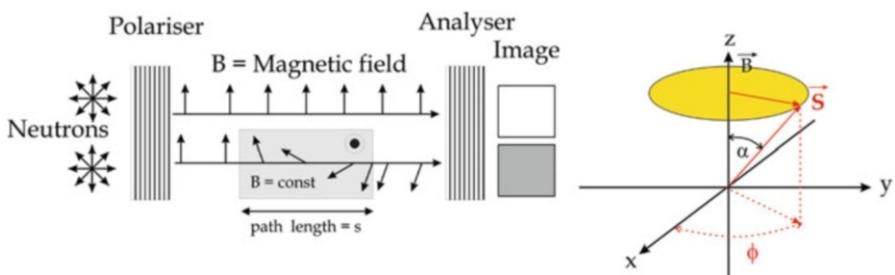
If a neutron with spin  $\mathbf{S}$  enters a magnetic field and  $\mathbf{S}$  and  $\mathbf{B}$  are not parallel to each other  $\mathbf{S}$  rotates around  $\mathbf{B}$  with the Larmor frequency  $\omega_L = \gamma \cdot B$  (see Fig. 22).

Therefore, the (total) rotation angle  $\phi$  is a measure of the strength of the magnetic field  $B$ . From  $\omega_L = \frac{\phi}{t}$  one gets  $\phi = \omega \cdot t$  and one can write  $\phi = \omega_L \cdot t = \gamma \cdot B \cdot t$ . 't' is the time the neutron needs to pass the B-field (see Fig. 31), with  $t = \frac{s}{v}$  the rotation angle becomes  $\phi = \gamma \cdot B \cdot \frac{s}{v}$ , i.e.  $\phi$  is directly proportional to  $B$ ,  $\phi \propto \phi(B)$ ,

$$\left( \phi = \frac{\gamma}{v} \cdot \int_{\text{path}} B(x, y, z) \cdot ds \right).$$



**Fig. 30** Magnetic field of a coil, the magnetic field along the red line inside of the coil is constant and outside it decreases rapidly



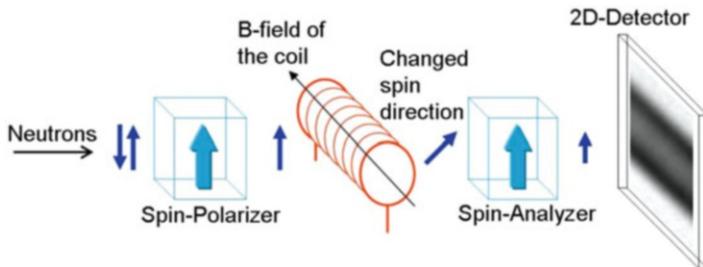
**Fig. 31** Layout of a standard polarized neutron imaging set-up, right:  $\mathbf{S}$  rotates around  $\mathbf{B}$  with the Larmor frequency  $\omega_L = \gamma \cdot B$

Considering the magnetic field in a coil, one remembers that  $B$  is homogeneous. The shape is cylindrical, and special projections are a circle (rod axis is parallel to neutron beam or a rectangle if the rod axis is perpendicular to neutron beam (see Fig. 32).

In order to illustrate the procedure of magnetic field determination using the Radon transform, consider for convenience two simple cross sections that are often used as samples, a cylinder (slice, plate) and cuboid. In the case of a cuboid sample, the fringe pattern due to a constant  $B$ -field looks like that shown in Fig. 33.

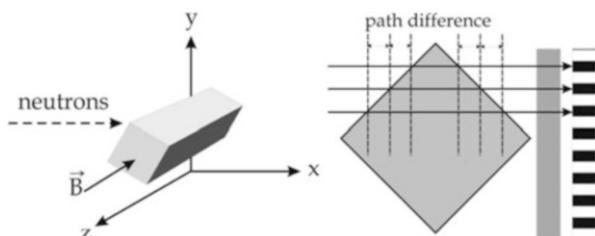
The neutron path length  $s$  increases linearly to the middle of the cuboid in order to decrease in the same manner to the bottom of the sample. This shape creates equidistant path differences causing an equidistant fringe pattern (see Figs. 33, 34 and 35).

Different from the cuboid sample, the corresponding path lengths through a cylindrically or disc shaped sample increase as  $s(y) = 2 \cdot \sqrt{r^2 - (r - y)^2}$  and do not form an equidistant fringe pattern of homogeneous magnetic field.

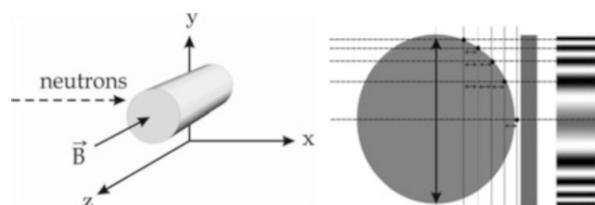


**Fig. 32** Layout of a standard polarized neutron imaging set-up

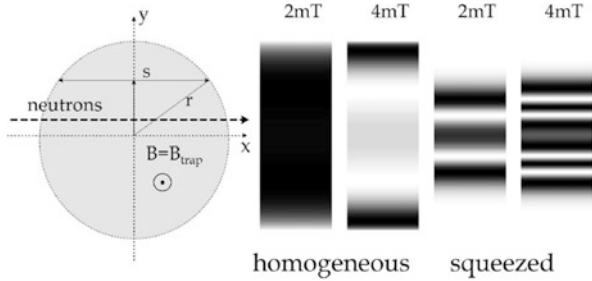
**Fig. 33** Sample position and fringe pattern



**Fig. 34** Sample orientation with respect to the neutron beam and corresponding fringe pattern



**Fig. 35** Image of a calculated fringe pattern due to polarized neutron spin interaction with a homogeneous magnetic field



The calculation of a path integral (see Eq. 76) along a line through  $B$  yields the weighted sum of  $B(x, y, z)$  along the path through  $B$ . This can be done applying the Radon transform to  $B = B(x, y, z)$  (Herman 1980; Deans 1983). Following the definition in (Deans 1983), the 2D-Radon transform is defined as the mapping of an arbitrary (finite) function  $f$  defined in the  $(x, y)$  plane (domain) by the projection or line integral of  $f$  along all possible lines  $L$ .

$$\mathcal{RT}\{f(x, y)\} = \int_L f(x, y) ds \quad (92)$$

$ds$  is the differential path length along  $L$ . Applying this to  $B = B(x, y, z)$ , it is subdivided into  $y$ -slices, and one gets with (Eq. 92) the Radon transform of the  $y$ th slice (Fig. 36)

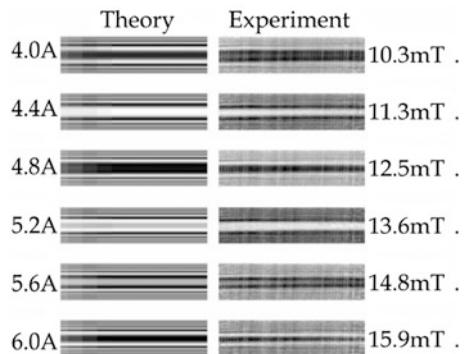
$$\mathcal{RT}_\theta\{B_y(x, z)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B(x, z) \cdot \delta(t - x \cdot \cos(\theta) - z \cdot \sin(\theta)) \cdot dx \cdot dz \quad (93)$$

One gets for the sample orientations (see Fig. 33),  $\theta = 0^\circ$  and  $90^\circ$ ,

$$\begin{aligned} \mathcal{RT}_0\{B\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B(x, z) \cdot \delta(p - x) \cdot dx \cdot dz \\ \mathcal{RT}_{90}\{B\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B(x, z) \cdot \delta(p - z) \cdot dy \cdot dz \end{aligned} \quad (94)$$

Both Radon transforms (Eq. 94) yield the path integrals for the particular orientation of the sample with respect to the neutron beam. Thus, with (Eq. 94) and known path length  $s$ , one can calculate  $\phi$ . The result of such a 2D radiography (with or without an external magnetic field) is a two-dimensional depolarization matrix, which is used to determine the amount and the shape of the expelled and trapped/pinned magnetic field. For both sample orientations, the intensities in the  $(x, y)$ -plane and  $(y, z)$ -plane become

**Fig. 36** Comparison: Calculated fringe pattern and experimental results. The current of a coil was increased up to 6 Ampere and for each current (magnetic field) an image was recorded.  
(Diameter of the coil 5 mm)



$$I(x, y) = \underbrace{I_0 \cdot T \cdot \exp \left( - \int_{\text{path}} \sum(s) ds \right)}_{I_{\text{att}}(x, z)} \cdot \underbrace{\frac{1}{2} (1 + \cos \phi(x, y))}_{I_{\text{spin}}(x, y)} \quad (95)$$

$$I(y, z) = \underbrace{I_0 \cdot T \cdot \exp \left( - \int_{\text{path}} \sum(s) ds \right)}_{I_{\text{att}}(y, z)} \cdot \underbrace{\frac{1}{2} (1 + \cos \phi(y, z))}_{I_{\text{spin}}(y, z)} \quad (96)$$

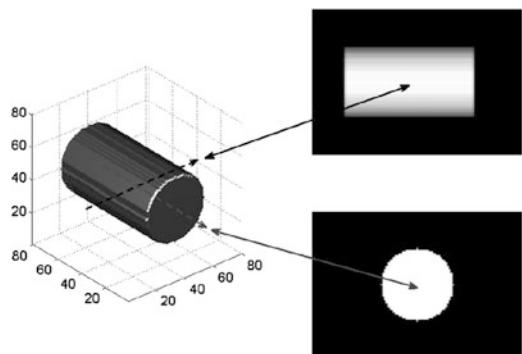
$I_{\text{att}}(x, z)$  and  $I_{\text{att}}(y, z)$  are the attenuated intensities in the x-direction and z-direction, measured in the  $(y, z)$  detector plane and  $(x, y)$  detector plane, respectively,  $\phi$  is the spin rotation angle when the neutron travels along the x-axis ( $\phi(x, y)$ ) or along the z-axis ( $\phi(y, z)$ ). In both orientations, the path through  $B = B(x, y, z)$  ‘y’ denotes the height-dependence of the ray. The cross section for  $\theta = 0^\circ$  is a rectangle, for  $\theta = 90^\circ$  it is a circle (Fig. 37).

For all calculations mentioned above, the magnetic field  $B = B(x, y, z)$  must either be reconstructed from projections (quite difficult) or one assumes 3D models of magnetic fields that might represent the trapped field. It will be shown that for a continuous trapped field, as occurs in superconductivity, the following procedure leads to good results.

- Create a model of  $B = B(x, y, z)$
- Calculate the Radon transform  $\mathcal{RT}$  of slices of  $B(x, y, z)$
- Transform  $\mathcal{RT}\{B(x, y, z)\}$  into 2D-spin depolarized images (modulo  $2\pi$ )
- Compare the calculated image with the experimental one
- Verify magnetic interaction  $\Leftrightarrow$  spin up–spin down images

In order to find a ‘good’ model for a trapped field, e.g. in a superconductor, one must take into account some boundary conditions. In superconductor type I critical

**Fig. 37** Two projections of a cylindrical sample, orientated parallel ( $0^\circ$ ) and perpendicular ( $90^\circ$ ) to the neutron beam, assuming small absorption



fields  $B_c(T)$  at 0 K are usually  $\leq 0.1$  Tesla, which decreases to zero when the temperature  $T$  approaches  $T_c$ . (Poole et al. 2007)

$$B_c \approx B_c(0) \cdot \left[ 1 - \left( \frac{T}{T_c} \right)^2 \right] \quad (97)$$

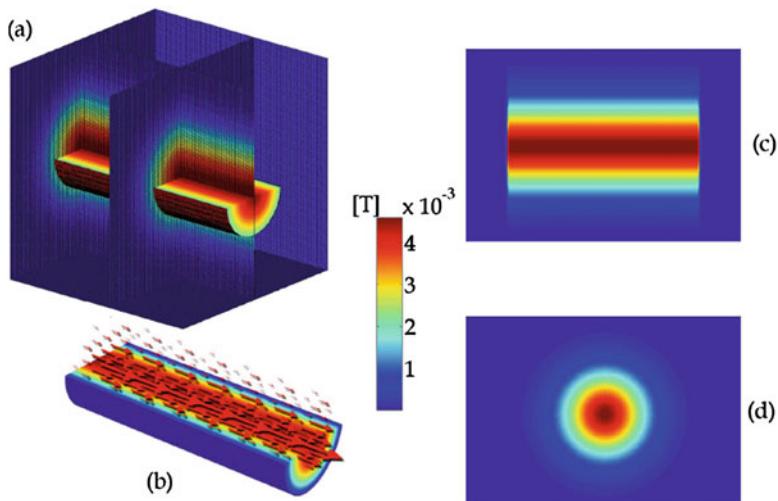
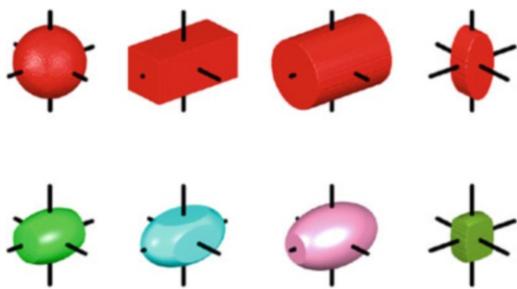
Therefore, magnetic (stray) fields outside the sample will be low, decreasing as  $B \sim \frac{1}{r}$ , if  $r$  is the distance from the centre of the sample. Thus, for a small trapped field, the magnetic field outside the sample can be neglected. The modelling procedure starts with the shape of the sample because the trapped field itself will first have the shape of the sample and is supposed to be zero outside the sample. A 3D volume can be realized, e.g. with MATLAB, see Fig. 38. The shape of the field can be approximated, e.g. by 2D or 3D Gaussian functions. From experiment, one gets 2D images that are the result of depolarized neutrons due to a magnetic field mapped on a 2D-position sensitive detector. With this method only the amount of a trapped field can be modelled (Fig. 39).

## Experiments with Polarized Neutron Imaging

### Example: Superconductor Type I

The instrument PONTO – POlarized Neutron TOmography @ BER II reactor (Berlin, Germany, operated by the University of Applied Sciences Beuth Hochschule Berlin, closed in 2017), is shown in Fig. 40. A graphite (C-) monochromator in the shielding reflects neutrons with  $\lambda = 0.32$  nm towards an optical bench, which could be rotated around the C-monochromator, a vertical and horizontal collimator restricted the beam in both directions to  $6'$  ( $1.745 \cdot 10^{-3}$  rad) or supplying a  $L/D = 573$ . Behind a spin polarizer the spin passed a guide field to the sample. The sample itself was – in the case of experiments with superconductors – in a cryostat (not shown in Fig. 40) between two Helmholtz coils that generated maximum external magnetic fields  $B_{ext}$  up to 60 mT. The 2D-detector had a pixel size of

**Fig. 38** Possible shapes of samples (top row) and amount of possible trapped fields (bottom row), both calculated with MATLAB. Note, only the strength (amount) of  $B$  can be given



**Fig. 39** 3D image of a trapped field in a  $Pb$  sample (Treimer et al. 2012b)



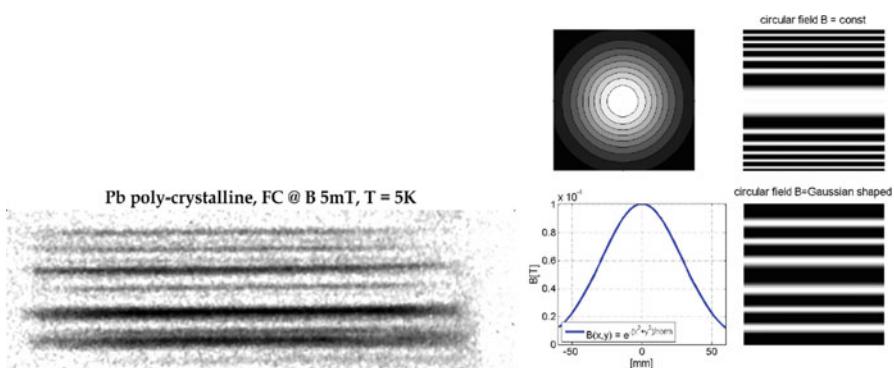
**Fig. 40** Layout and instrument PONTO at the BER II reactor, HZB (see text). (Treimer et al. 2014)

13.5  $\mu\text{m}$ , the spatial resolution was 55  $\mu\text{m}$  in the case of unpolarized neutrons and 120  $\mu\text{m}$  for polarized neutrons. The worse resolution for polarized neutron imaging was first affected by the diameter of the cryostat (increased distance to the detector) and then more by the stray field of the spin analyser. In order to get a nearly ‘zero-B field’ at the sample position, the cryostat and the Helmholtz coils must be put at a distance of approximately 16 cm apart from the analyser. An experimental example for the determination of a trapped field in a cylindrically shaped superconductor type I is given below.

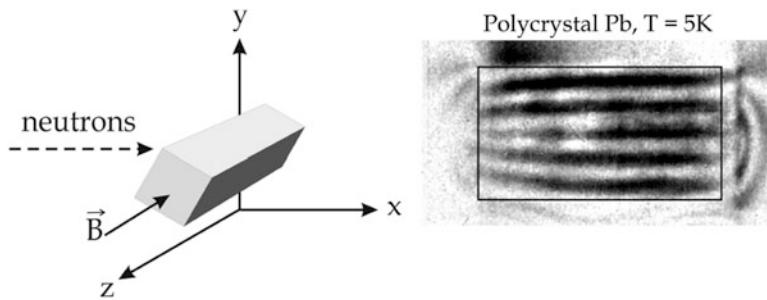
A polycrystalline Pb cylinder (12 mm in diameter, length = 30 mm) was cooled down from  $T = 8 \text{ K}$  to  $T = 5 \text{ K}$  in the presence of an external field  $B_{\text{ext}} = 5 \text{ mT}$ . Then at  $T = 5 \text{ K}$ ,  $B_{\text{ext}}$  was switched off and a radiography with polarized neutrons was recorded (exposure time  $\sim 4 \text{ h}$ ). In the case of a perfect Meißner state (expulsion of  $B$ ), the Pb sample should not show a trapped field, i.e. be visible; however, one recognizes field trapping as shown in Fig. 41. It was expected that a trapped field in a cylindrical shaped sample will show a fringe pattern as a coil (see Fig. 34); however, experimentally it could be proved a different trapping, as shown in Figs. 39 and 41. This is one of several achievements of imaging with polarized neutrons.

Superconductivity is a thermodynamic phase, and therefore an interesting parameter for the study of its behaviour for  $T < T_c$ . In the following, an example is given for imaging the temperature dependence of a trapped field in a superconducting Pb cuboid shaped sample. The temperature dependence on the trapped field was investigated in the Meissner phase (Treimer et al. 2012a, b). Therefore, the questions to be answered are (Fig. 42):

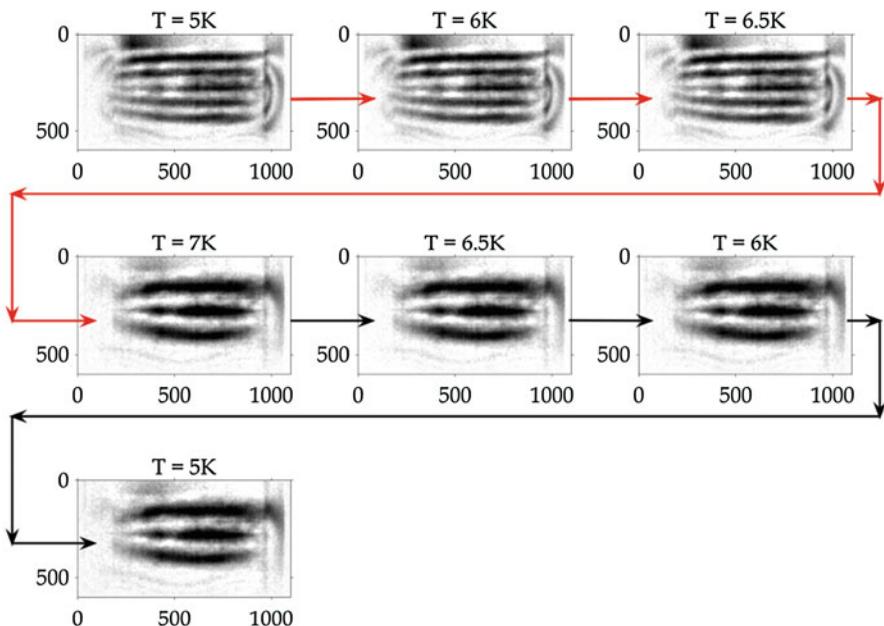
- Is  $B$  really completely expelled for FC from  $T > T_c \rightarrow T < T_c$ ?
- What does a trapped field look like?
- What happens with a trapped flux if  $T_1 < T_2$ :  $T_1 \rightarrow T_2 < T_c$ ?
- Role of the shape/size of the sample?



**Fig. 41** Left: Radiography of superconducting Pb at 5 K (FC at  $B = 5 \text{ mT}$ ), note the fringes which are nearly equidistant. Right: calculated images of homogeneous and squeezed trapped field



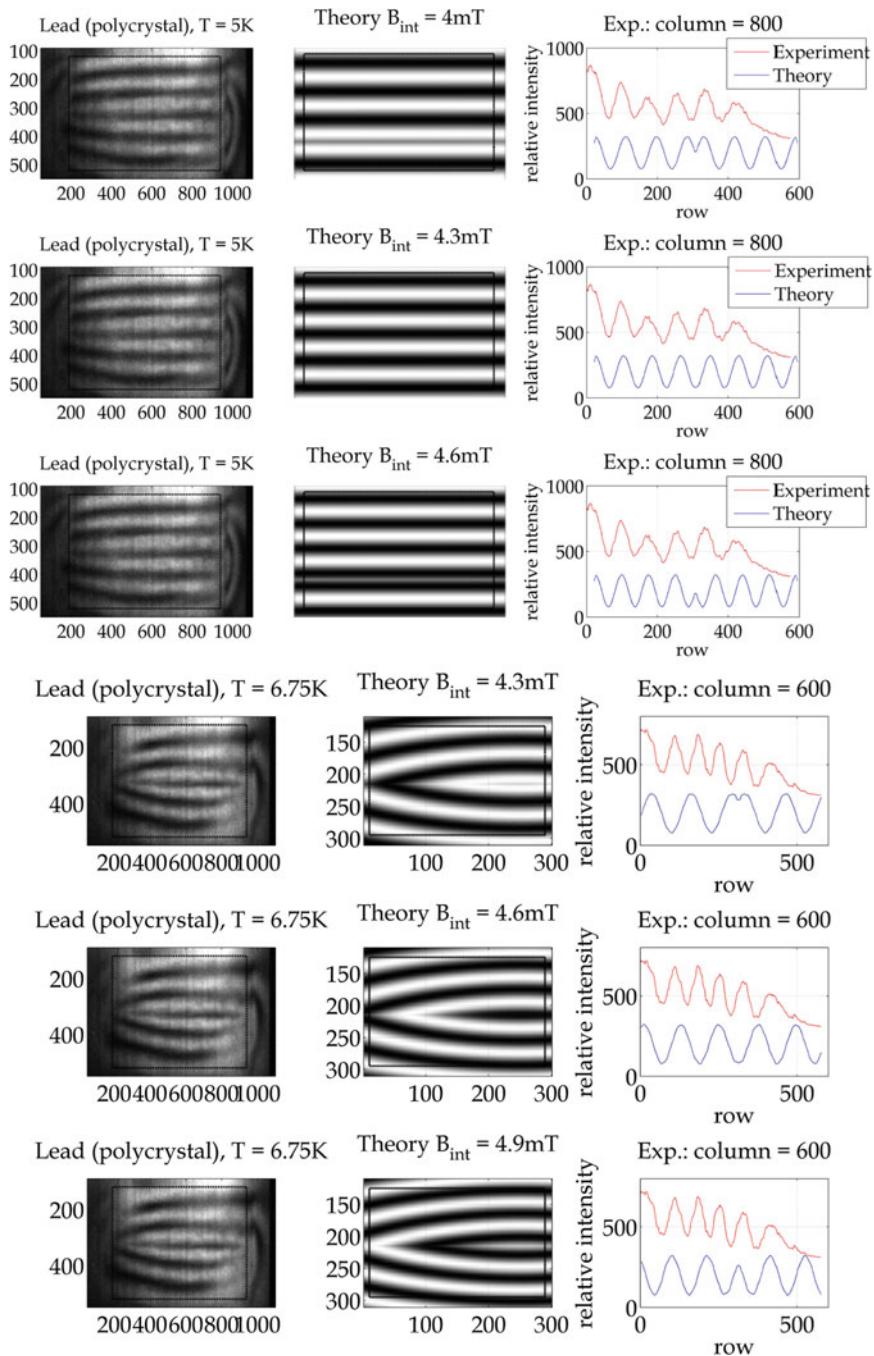
**Fig. 42** Polycrystalline  $Pb$  cuboid, cross section  $12\text{ mm} \times 12\text{ mm}$ , length =  $30\text{ mm}$ , external  $B$ -field parallel to rod axis, FC at  $B = 5\text{ mT}$ , observed fringe pattern appears equidistant  $\leftrightarrow$  homogeneous field trap

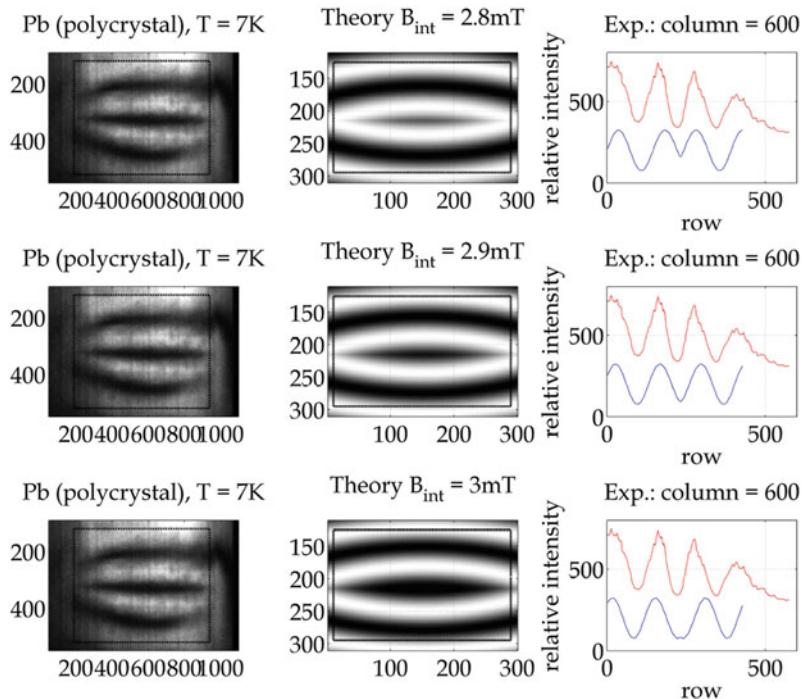


**Fig. 43** Study of trapped fields in  $Pb$  in the Meißner phase (see text)

The first results have shown an equidistant fringe pattern which can – for the cuboid sample – originate only by a homogeneous trapped magnetic field. In order to investigate this behaviour, each image in Fig. 43 was modelled with a trapped field (see e.g. Fig. 38) and compared with the corresponding images as shown in Fig. 44, (three images).

Left: The  $Pb$  sample was cooled from  $T = 8\text{ K}$  down to  $T = 5\text{ K}$  in the presence of an external field  $B_{\text{ext}} = 5\text{ mT}$ , and after switching off  $B_{\text{ext}}$  for each temperature  $T = 5\text{ K} \dots T = 7\text{ K}$ , a radiography with polarized neutron was done (see Fig. 43). The critical temperature  $T_c(Pb)$  is  $7.2\text{ K}$ , so the images were recorded still below  $T_c$ .

**Fig. 44** (continued)



**Fig. 44** Experimental and calculated images of trapped fields in  $Pb$  sample

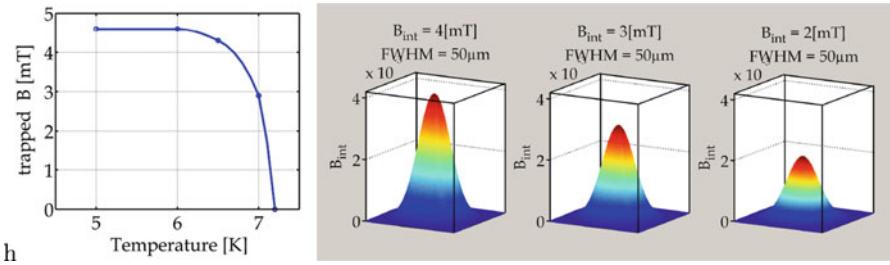
One recognises a quite constant trapped field until  $T = 6.5$  K. Between  $T = 6.5$  K and  $T = 7$  K a change in the pattern occurred, i.e. the number of fringes reduced. Cooling the sample down again to  $T = 5$  K did not change the trapped field

The results from Figs. 44 and 45 underline the importance of polarized neutron imaging. With this method one can ‘look inside’ of massive samples ( $\text{cm}^3$  large) and visualize magnetism with the spatial resolution of  $\approx 120 \mu\text{m}$ . Niobium samples have also been investigated with this method (Aull et al. 2012; Treimer 2014).

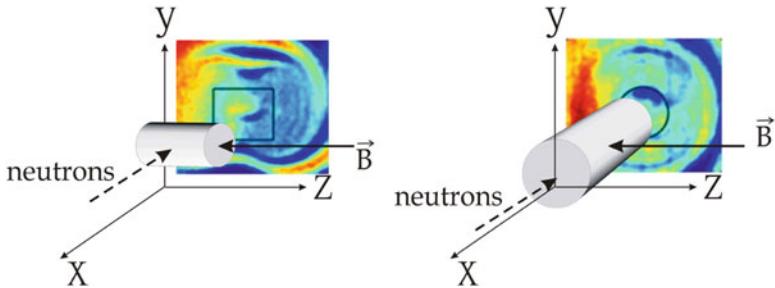
### Example: Superconductor Type II

The investigation of superconductor type II is different from type I because in the superconducting phase there are two different critical fields  $H_{c1}$  and  $H_{c2}$ . The range between  $H = 0$  and  $H = H_{c1}$  is called also the Meissner phase, showing the same behaviour as superconductor type I, i.e. total exclusion or expulsion of external magnetic fields, i.e. perfect diamagnetism. The range between  $H_{c1}$  and  $H_{c2}$  is a mixed state where an external magnetic field  $H_{c2} > H > H_{c1}$  can enter the sample by means of so-called Abrikosov flux lines (vortices) (Poole et al. 2007; Annert 2013).

As an example for a superconductor type II, the study of T-dependencies of trapped fields in  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$  with  $x = 0.09$  is given.  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$  is an antiferromagnet, a high- $T_c$  superconductor and was discovered about three decades



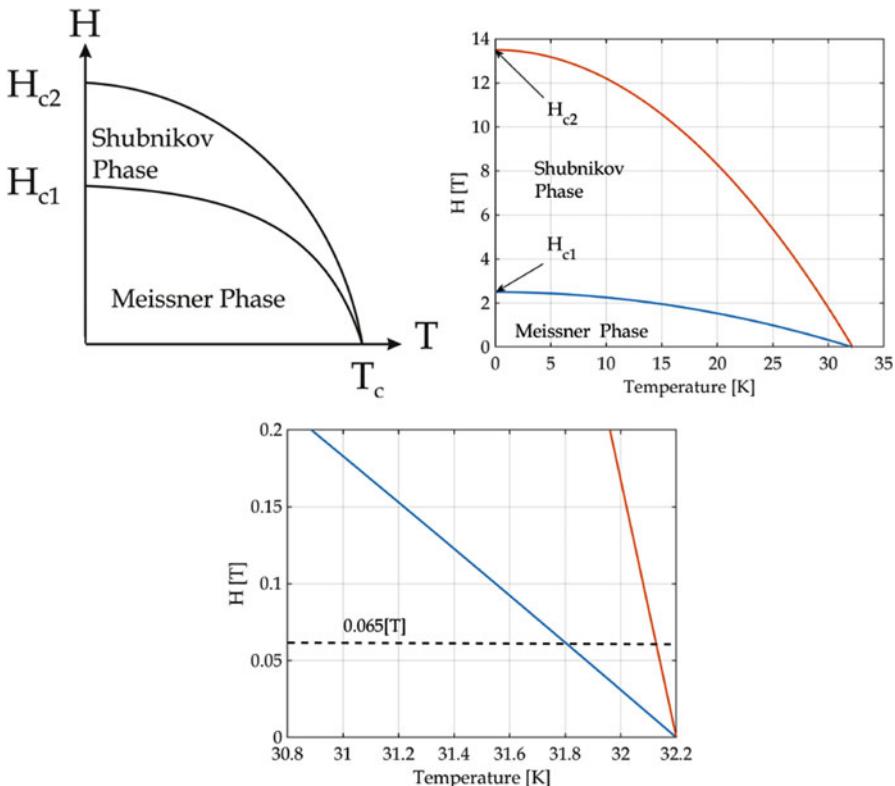
**Fig. 45** Left graph: temperature dependence of trapped field, right: calculated shape of trapped field as a function of temperature  $T$ , see text



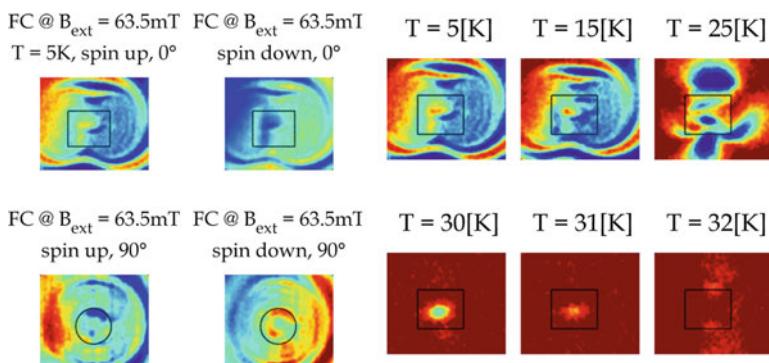
**Fig. 46** Sample orientation with respect to external  $B$  field and incident neutrons. Note the stray field around the sample (indicated by a black rectangular and circle, respectively); red color = spin up detection, blue color = spin down detection, exposure time for each image was 4 h)

ago (Müller and Bednorz 1987; Tarascon et al. 1987; Takagi et al 1989; Shi et al. 2014; Iavarone et al. 2014; Keimer et al. 1992). The critical fields are  $H_{c1} = 2.5$  T and  $H_{c2} = 13.5$  T, and the critical temperature  $T_c = 32.2$  K  $\pm 0.2$  K. Recent studies of this sample by polarized neutron imaging showed the temperature dependence of flux pinning and flux pinning as a function of magnetic field. (Dhiman et al. 2017).

Here, the method described above is used in order to determine that the trapped field is different from the one given in (Dhiman et al. 2017). In Fig. 46, two sample orientations with respect to the external magnetic field  $B$  and incident neutrons are shown. The measuring method as written above is the same: Field cooling (FC) from  $T > T_c = 32.5$  K in the presence of an external magnetic field  $B_{\text{ext}} = 63.5$  mT down to  $T = 5$  K. The magnetic field is parallel to the rod axis. Then the magnetic field is switched off and four images are recorded, for sample orientations  $0^\circ$  and  $90^\circ$  and for each spin up and spin down neutrons. As one can see in Fig. 47, to distinguish the Meissner phase from the Shubnikov phase, the applied field must be much larger than the availablely used 65 mT. Therefore, measurements were performed at  $T = 5$  K, 15 K, 25 K and  $T = 30$  K, 31 K and 32 K, still below  $T_c = 32.2$  K. The behaviour of the trapped field is shown in Fig. 48. Different from what one can expect from Fig. 47, one measures large changes of trapped field between  $T = 15$  K and  $T = 25$  K and also between  $T = 25$  K and  $T = 32$  K. For an external magnetic field of

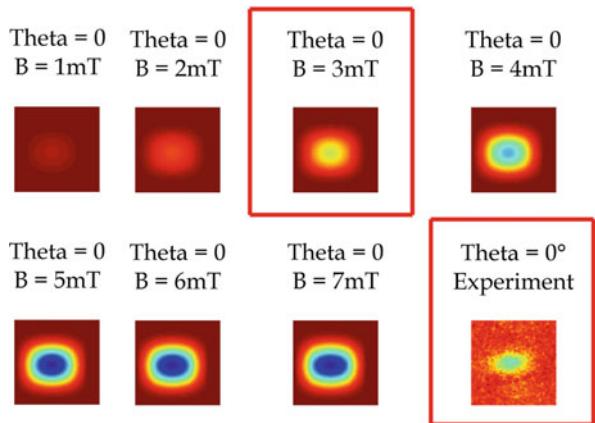


**Fig. 47** Left figure: T-H phase diagrams of superconductors type II, middle and right figure: the T-H phase diagram of  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$  using  $H_{c1} = 2.5$  T and  $H_{c2} = 13.5$  T



**Fig. 48** Left four images: Spin up and spin down images after FC at  $B = 63.5$  mT for  $T = 0$  5 K for sample orientations  $0^\circ$  and  $90^\circ$ , right six images: Warming up from  $T = 5$  K up to  $T = 32$  K, black rectangle is the shape of the sample; red color = spin up detection, blue color = spin down detection, exposure time for each image is 2 h)

**Fig. 49** Calculated neutron images for trapped fields  $B = 1 \text{ mT} \dots 7 \text{ mT}$ , for  $T = 30 \text{ K}$ , the best agreement with measured image is found for a trapped field  $B = 3 \text{ mT}$



$B = 63.5 \text{ mT}$ , the Meissner phase should be present until  $31.8 \text{ K}$ , but a (dramatic) change is already observed at  $T = 25 \text{ K}$ . The trapped field decreases from  $B = 25(5) \text{ mT}$  to  $B = 3 \text{ mT}$  at  $T = 30 \text{ K}$ , where the sample should still be in the Meissner phase. The decrease and absence of trapped field is proved by the lack of stray field in images at  $T = 30\text{--}32 \text{ K}$ . The existence of Abrikosov flux lines (diameter approximately  $1 \mu\text{m}$ ) could not be observed owing to lack of necessary spatial resolution and to the small number of lines which are proportional to  $B$ .

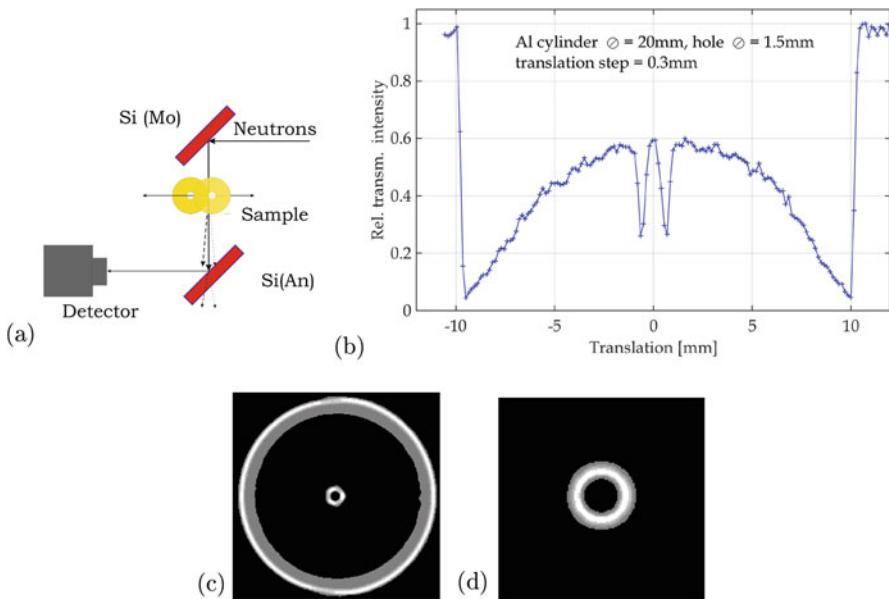
Figure 49 shows the comparison of the measured and calculated neutron images for  $T = 30 \text{ K}$  (see Fig. 48). The amount of the trapped field was assumed to have a 3D Gaussian shape defined by  $\sigma_x$ ,  $\sigma_y$ ,  $\sigma_z$ , and the best agreement was found for  $\sigma_x = 34 \text{ mm}$ ,  $\sigma_y = 17 \text{ mm}$  and  $\sigma_z = 10 \text{ mm}$  (see Fig. 46).

## Phase-Based Imaging Signals

### Some Historical Aspects and Basics

There are a number of examples in neutron imaging where absorption or attenuation based interactions stemming from small defects or material boundaries are hard to distinguish from surrounding matter. The reasons are either the defects are too small, and/or their attenuation coefficients are very similar to the surrounding matter.

Attempts to overcome this disadvantage were done in 1996, where refraction contrast was investigated in the frame of neutron optics and dynamical neutron scattering (Schaper 1996). In this diploma work, it could be demonstrated that even (for neutron) transparent structures could be detected if phase effects such as refraction could be clearly separated from surrounding contrast. This could be realized by using a double crystal diffractometer (see Fig. 50) and its  $\mu\text{rad}$  angular resolution (due to a FWHM of the analyser rocking curve less than  $\mu\text{rad}$ ). An Al cylinder ( $\varnothing = 20 \text{ mm}$ ) with a central hole ( $\varnothing = 1.5 \text{ mm}$ ) placed between



**Fig. 50** (a) Set up double crystal diffractometer, (b) refraction by a cylindrical shaped Al sample with a hole is translated stepwise perpendicular through the neutron beam and for each translation position the transmitted intensity is measured. Note its special behaviour, mainly dominated by refraction and not by absorption, (c) 2D reconstruction of cylinder with hole as (d) details reconstructed from translation data. (Schaper 1996, Treimer 1998, see text)

monochromator (Si(Mo)) and analyser crystal (Si(An)) was translated perpendicular to the neutron beam (width = 0.35 mm) and the transmitted intensity recorded as a function of translation (layout see Fig. 50). This translation curve is characteristic for phase-based imaging if refraction comes into play. A corresponding scan recording absorption only would look entirely different (DCD type, see Fig. 50). Because ultra small angle scattering and refraction change the flight direction of neutrons without loss of beam intensity, the use of a double crystal diffractometer for neutron imaging started in the beginning of the 1990s (Treimer et al. 1991, 1998; Maass et al. 1992; Treimer et al 1997; Treimer and Feye-Treimer 1998).

Further experimental improvements have been realized by four diploma theses (Kraft 1996; Schaper 1996; Ernst 1997; Herzig 1997) setting the basis for following developments. The underlying idea for all tomographic reconstruction is to observe and distinguish the characteristic interaction (absorption, refraction, small angle scattering) and its action on neutrons.

If a neutron enters a sample (e.g. the Al cylinder as described above), refraction occurs at the lateral surface of the cylinder and refracted neutrons are not reflected by the analyser crystal, which is set parallel (The reflecting atomic planes must be parallel to each other) to the monochromator crystal (Si(Mo) and Si(An) in Fig. 50), yielding a strong decrease of intensity in the detector; the same happens at the hole edges. The decrease of intensity due to refraction happens simultaneously with

absorption, which is low at the edges of the cylinder and increases with increasing path length in the sample. Pure absorption would show an entirely different graph than observed. The sample was scanned step-wise perpendicular to the beam and the transmitted intensity measured for each step. This procedure can be performed in different ways: First, the analyser is kept in its peak position (maximum reflected intensity) and any change of the beam direction between the crystals will cause a decrease of the intensity reflected by the analyser. In this case, one gets no information about in which direction the beam was deviated.

The second method is to keep the analyser at one of its slopes of the rocking curve at the half-maximum-position. Then a deviation will cause a decrease but also an increase of the reflected intensity, depending on in which direction the beam was deviated.

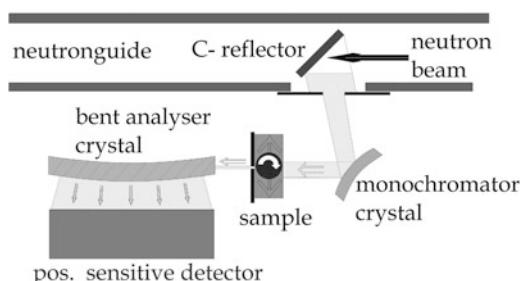
Considering an arbitrary sample, with this technique one can get information about the sum of deviations; however, it is limited (in one direction) to the maximum deviation of half of the FWHM. In the case of refraction, absorption and (ultra) small angle scattering must also be taken into account, which requires the determination of all transmitted neutrons, i.e. the measurement of the total rocking curve.

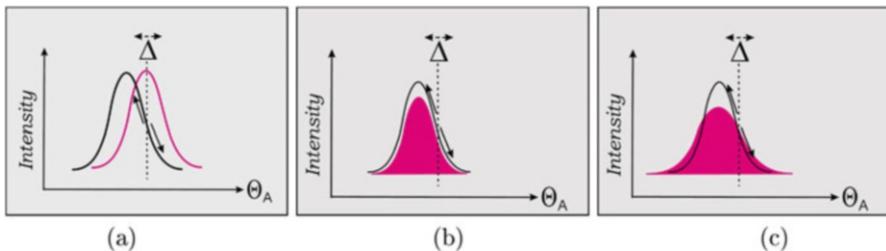
Another approach to use phase-based interaction as an imaging signal is (ultra) small angle neutron scattering ((U)SANS). Structures of the sizes 1 nm up to 1  $\mu\text{m}$  are investigated by SANS, measuring the scattering curve from structures in the sample. For imaging, the broadening of the scattering curves is used as the imaging signal. The proof of principle was performed with a double crystal diffractometer, which allowed the simultaneous measurement of attenuation, refraction and small angle scattering (Figs. 51 and 52).

Tomographic measurements with a double diffractometer take quite a long measuring time and results are seen as proof of principle. Later on, refraction and small angle scattering techniques were methodically improved (Treimer et al. 2003; Strobl et al. 2004a) and could be successfully applied to polarized neutrons (Strobl et al. 2007) and neutron tomography using neutron grating interferometry (Gruenzweig 2009; Gruenzweig et al. 2008a; Strobl et al. 2008a) (s. below).

When using phase-based signals for neutron imaging, some features must be explained in more detail. Refraction is well known from light optics, and it describes the change of propagation direction, if light enters from one (transparent) medium into another one (e.g. from air to glass). This change is caused by different but

**Fig. 51** With a special double crystal diffractometer all three interactions (absorption, refraction and small angle scattering) can be measured simultaneously





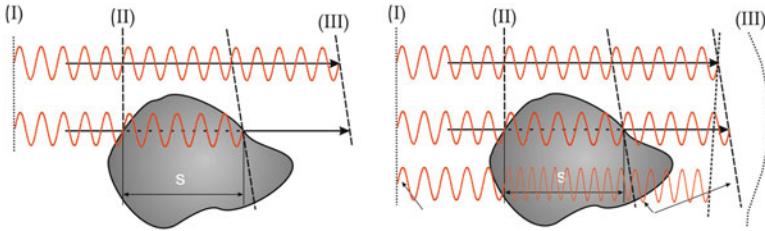
**Fig. 52** From left to right: (a) Refraction causes a shift of the rocking curve, (b) absorption diminishes scattered intensity (smaller rocking curve) and (c) small angle causes a spreading of scattered intensity, i.e. broadening of the rocking curve

coherent parts of the corresponding wave entering matter. A finite lateral (transversal) and longitudinal coherent length of the matter-wave is essential for this change. A classical particle would not – could not – act in this or in a similar way because it has no finite coherence length; its coherence length is virtually zero. When assigning neutron wave character, wave considerations always involve coherence, phase and amplitude calculations. The calculation of intensity involves phases and amplitudes of all points of the wave, which must be added and superimposed to build up the resulting amplitude and thus the intensity in the point of observation. Neutrons will be treated here as matter waves, and therefore all considerations concerning their interaction with matter are based on principles known from light optics. De Broglie has shown that the classical momentum  $p = m.v$  ( $m$  is the mass of the neutron and  $v$  its (particle) velocity) and wavelength  $\lambda$  of the corresponding matter-wavelength are connected with each other by the Planck constant  $h$ , as  $\lambda = h/p$ . Thus, the change from the particle aspect of the neutron to neutron optics can be seen to be quite simple, but dealing with optical parameters excludes particle aspects and vice versa.

All phase effects can be classified as coherent scattering and the important entity is the index of refraction  $n$ , which consists of a real and an imaginary part (see section “[Neutron Properties: Index of Refraction](#)”). The real part is responsible for phase-based interactions, i.e. elastic scattering (refraction, small angle scattering), whereas the imaginary part determines the attenuation (absorption) of neutrons by the nuclei (see below). All interactions of neutrons based on phase and phase shifts require quantum mechanics descriptions.

A neutron can be assigned to a matter wave and be treated as a classical wave. Travelling first in free space and then entering a homogeneous medium it experiences a phase shift with respect to the ‘free space wave’ and possibly a decrease of the amplitude of the wave. This interaction is called refraction. If the neutron wave hits a particle or structure that is embedded in homogeneous medium (matrix) and has sizes less than the coherence width of the wave, small angle scattering occurs. Both interactions are the basis of phase contrast neutron imaging, which is performed by different techniques (Fig. 53).

To introduce refraction as an imaging signal, one has to consider two (or more) points of a wave emerging from a point source.



**Fig. 53** Depending on the lateral width of the coherent wave front (I), a change of the wave front (III) causes refraction (left image) or small angle scattering (right image)

From optic lectures it is well known that two (or more) points of a wave emerging from a point source can be regarded to be coherent to each other if there is a constant phase relationship between them or if there is a given function that describes the evolution of the phase in space or time. It is further known that neutrons can be seen as particles as well as waves, and therefore the distinction between incoherent signals and coherent ones simplifies the discussion about different interactions in neutron radiography and tomography. The fundamental entity for all interactions (coherent, incoherent) is the index of refraction  $n$ , given by

$$n = 1 - \delta + i\beta \quad (98)$$

(Equation 98) can be written as (Rauch and Werner 2000)

$$n = \frac{K}{k} \sim 1 - \lambda^2 \frac{N}{2\pi} \sqrt{b_c^2 - \left(\frac{\sigma_t}{2\lambda}\right)^2 + i \frac{\sigma_t N \lambda}{4\pi}} \quad (99)$$

Here,  $K$  is the wave vector inside,  $k$  the wave vector outside of the material,  $\lambda$  the vacuum wavelength,  $N$  atomic density,  $b_c$  coherent scattering length of the material (for neutrons) and  $\sigma_t = \sigma_a + \sigma_{incoh}$  is the sum of the absorption and incoherent cross section. The material is seen as a volume having the mean interaction potential  $\bar{V}$  which is connected with  $n$  as

$$n = \sqrt{1 - \frac{\bar{V}}{E}} \quad (100)$$

The mean potential  $\bar{V}$  or optical potential is given by (Rauch and Werner 2000)

$$\bar{V} = \frac{2\pi\hbar^2}{m} b_c N \quad (101)$$

The mean optical potential can easily be calculated if  $b_c$  is known. The real part in (Eq. 99) describes the phase-based interaction, if  $a_t = \sigma_a + \sigma_{incoh}$ ,  $\sigma_t/2\lambda$  is much smaller than 1, and one can define a  $\delta$  as

$$\delta = \lambda^2 \frac{Nb_c}{2} \pi, \quad (102)$$

The imaginary part is then given by  $\beta$

$$\beta = \frac{\sigma_t N \lambda}{4\pi} \quad (103)$$

and it is responsible for the absorption part of the index of refraction. It is apparent that  $\delta$  corresponds with the wave-nature and  $\beta$  with the particle-like nature of the neutron. For simplicity, one considers monoenergetic neutrons that pass a layer of material having the mass thickness  $d$  and the density  $\rho$ , and the ratio of incident and transmitted intensity is then given

$$\frac{I}{I_0} = e^{-\frac{\mu}{\rho}d} \quad (104)$$

$\mu/\rho$  = mass attenuation coefficient. The mass thickness  $d$  is defined as the mass per unit area and is obtained by  $d = \rho \cdot t$ , with  $t$  = thickness of the layer. In the case of neutrons,  $\mu = \mu(E)$  depends on their energy (velocity) and on the nucleus (isotope). The absorption cannot be calculated from first principles of quantum mechanics as can be done in the case of x-rays. This is due to the different kind of interaction of neutrons and x-rays: neutrons interact with the nucleus of the atom, X-rays with the electrons surrounding the nucleus. Thus, X-ray absorption depends on the photon energy. To distinguish different structures in an object by (neutron and X-ray) imaging, certain differences of measured 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

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad (105)$$

where the particular intensities  $I_{\max}$  and  $I_{\min}$  are determined in adjacent pixels ( $n \times n$  matrix,  $n \geq 3$ ) of a 2D detector or film. The contrast  $C$  is also determined by the kind of interaction, i.e. absorption, elastic and inelastic, coherent and incoherent scattering, spin interaction or excitation. Attenuation involves both, absorption and scattering. Therefore, attenuation of a neutron 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 can still be visualized by phase contrast. In the case of absorption, the interaction of radiation with matter is entirely incoherent, but in the case of phase contrast both coherent interactions of the incident beam and of the transmitted beam have to be taken into account. Therefore, it is very useful to separate attenuation contrast into an incoherent and coherent part. A ray that traverses an object often experiences 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).

In the case of neutron radiography and tomography, a sample, with shape geometrically defined in a bounded region, is scanned by pencil rays that traverse the sample, where  $\mu(x, y)$  characterizes the position dependent absorption of the sample by (mono-energetic) neutrons (see section “[Basics of Neutron Radiography and Tomography](#)”). If  $\mu = \mu(x, y)$  denotes the two-dimensional function that describes the index of absorption within a defined slice in the  $(x, y)$  plane,  $\mu(x, y)$  can be represented by the following path-integral

$$\mu(x, y) = -\frac{4\pi}{\lambda} \int_{\text{path}} \beta(s) ds \quad (106)$$

where  $s$  defines the path along a straight line over  $\mu(x, y)$ . The intensity ratio  $I/I_0$  is given by

$$\frac{I}{I_0} = \exp \left( - \int_{\text{path}} \mu(x, y) ds \right) \quad (107)$$

and a point of a so-called projection  $P_\theta(t)$  (see Fig. 7) is determined by

$$P_\theta(t) = \ln \frac{I_0}{I} = \int_{\text{path}} \mu(x, y) ds \quad (108)$$

One projection  $P_\theta(t)$  is a set of such line integrals covering the whole slice (sample). The mathematical solution to reconstruct a two dimensional function from its projections is well known (Herman 1980), and details are given in section “[Mathematics of Computerized Tomography](#).<sup>7</sup>” With the so-called method of filtered back projection (FBP), it is possible to reconstruct a slice of an object that was scanned over  $180^\circ$  by calculating the two dimensional Fourier transform  $M(u, v)$  of the image function  $\mu(x, y)$  from the one dimensional Fourier transform  $S(\omega, \alpha)$  of the projection  $P(s, \alpha)$ . The back transformation yields the two-dimensional function  $\mu(x, y)$ . If  $S_\theta(\omega)$  is the Fourier transform of  $P_\theta(t)$

$$\mu(x, y) \sim P_\theta(t) \otimes FT^{-1}\{|\omega|\} \quad (109)$$

In the case of absorption contrast,  $P_\theta(t)$  is the logarithmic ratio of the incident and transmitted intensity based on the particular absorption processes in the sample. It is measured, as indicated above, by scanning the sample with pencil rays for sample orientations  $\theta$  with  $0 < \theta < 180^\circ$ . Absorption as an imaging signal can be used in different ways, such as energy-dependent radiography and tomography, using the wavelength-dependent Bragg-edges of a cold neutron beam, real time radiography and tomography and CT with monochromatic neutrons. Problems arise if structures do not contribute significantly to the absorption signal. The question is how to get an image from these structures. Interactions other than absorption, such as phase-based interactions, have to be considered.

## Radiography and Tomography: Refraction

The coherent part of the index of refraction  $n$  in (Eq. 99) is responsible for the wave nature of the neutron and therefore also for its phase behaviour. Here, an approach is presented on how to use the phase as an imaging signal for radiography and tomography. Owing to coherent interaction with matter, the (coherent) partial waves form a wave front that changes its direction for  $z > d^2/\lambda$  ( $z$  = distance object–detector and  $d$  = diameter of the coherent illuminated object) (Treimer et al. 2005a, b; Strobl et al. 2004a). If the interface between two media in a sample is not perpendicular to the flight direction of the neutron, refraction occurs. This is especially true for tomography, where the incident flight direction changes from  $0^\circ$  to  $180^\circ$  (sample is rotated) during the data recording. The change of direction of an incident ray by refraction or small angles scattering is of the order of  $10^{-6}$  rad =  $\mu$  rad (for glancing angles  $< 5^\circ$ ) and therefore not observed in conventional tomography. This is due to the (extremely small) interaction potential of a nucleus with neutrons. Considering this interaction more closely, one has to distinguish between two cases, the one where the incident coherent wave front is larger and the other where it is smaller than the interface of the object. In the case of refraction, the coherent wave front is much smaller than the interface. A neutron (wave) passing through a sample is then many times deflected (zigzag path) by the structures in the sample (all of them smaller than coherence width) and leaves the sample with a different flight direction compared to the one in front of the sample. The index of refraction  $n$  becomes a 2-dimensional function of the coordinates  $(x, y)$  and is given as

$$n(x, y) = 1 - \frac{\lambda^2 N(x, y) b_c(x, y)}{2\pi} = 1 - \delta(x, y) \quad (110)$$

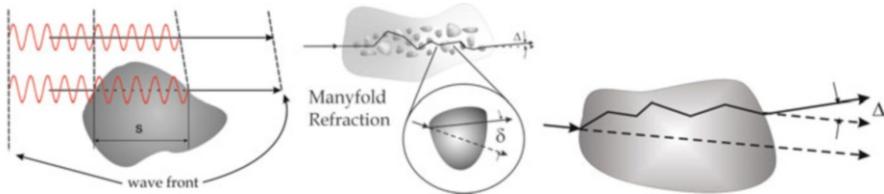
$N(x, y)$  = position dependent material density and  $b_c = b_c(x, y)$  = (mean) coherent scattering length. Refraction along a certain path can be described by a position dependent phase shift function  $\Delta\varphi(x, y)$ ,  $\delta = \delta(x, y)$  as (Treimer et al. 2003; Strobl et al. 2004a)

$$\Delta\varphi(x, y) = -k \int_{\text{path}} \delta(x, y) ds \quad (111)$$

Note that the  $\Delta$  is not the Laplace operator. The entities one has to determine are the refraction angles  $\Theta = \theta(\theta, t)$ , where  $\theta$  is orientation of the object related to the incident beam. One only measures remarkable  $\Theta$  if the interface is nearly parallel to the incident beam (small glancing angle); thus,  $\Theta$  can be approximated by (Fig. 54)

$$\Theta(\theta, t) = \frac{1}{k} \nabla_{\perp} (\Delta\varphi(x, y)) \quad (112)$$

A point of a projection  $P_\theta(t)$  can then be written as (see section “[Mathematics of Computerized Tomography](#)”)



**Fig. 54** Left figure: change of wave front by a structure larger than the coherent wave front, middle and right figure: multiple refraction results in a final deviation of the neutron  $\Delta$

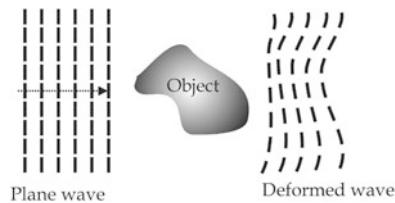
$$\begin{aligned}
 P_\theta(t) &= \int_0^t \Theta(x, y(x, \theta, t')) dt' \\
 &= \int_{\text{path}} \delta(x, y(x, \theta, t)) ds dx dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta_D(x \cos \theta + y \sin \theta - t) \delta(x, y)
 \end{aligned} \tag{113}$$

( $\delta_D$  = Dirac function). Here, one makes use of the assumption that the deviations of the ray from the incident direction are small ( $\mu\text{rad}$  that the path (lengths) through the sample, the same as without refraction effects. (This would not work with light, where  $\Theta \sim {}^\circ$ ). This means that in both cases the same pixels are hit. This holds for refraction and small angles scattering. The (numerical) integration is done along the path given by the same geometry as in the case of absorption. The set of all projections  $\{P_\theta(t)\}$  can now be used in the reconstruction algorithms described in section “[Mathematics of Computerized Tomography](#).“ It is worth mentioning that, using a double crystal diffractometer, magnetic refraction could also be measured and the magnetic field two-dimensionally reconstructed, setting the basis for further polarized neutron imaging (Strobl et al. 2007).

## Radiography and Tomography: Small Angle Scattering

Small angle scattering (SANS) is one of the most important tools in neutron physics to study structures in matter of the order of  $10^{-1}$  nm...  $10^3$  nm (Glatter and Kratky 1982). Especially in biology, SANS uses the difference of the scattering cross sections between structure and matrix. A very important fact – often ignored – is the knowledge of available coherence. Small angle scattering requires the coherent illumination of the whole structure to get full information from the scattering pattern. Thus, different from refraction, in small angle scattering (ultra small angle scattering) one has to consider whether a neutron wave traversing a structure has a coherent wave front of the order or larger than the structure (see Fig. 55). In this case, small angle scattering arises (or large angle scattering if diffraction is involved), which decreases the contrast in a conventional tomography but also can be used for neutron imaging (Treimer et al. 2003; Strobl et al. 2004a). As mentioned above, the size of

**Fig. 55** The incident wave front is larger than the object. The transmitted wave front is distorted owing to phase shifts in the object and causes small angle scattering



the incident lateral coherent wave front is a very important parameter to distinguish between different scattering processes that are used as imaging signals. If a neutron wave traverses structures that have sizes of the order or less of the coherent width of the wave front, the wave front is distorted after the passage through the sample. The phase front travelling through the object experiences different phase shifts due to different path lengths in the object. This behaviour – small angle scattering – is well known and used for investigations of structures having sizes between nanometer and micrometer.

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 (SAS) is an established method in X-ray and neutron physics and has a wide field of applications. Therefore, it was interesting to investigate the possibilities to use it as an imaging signal for tomography. In the case of SAS, absorption should not be the dominating interaction; however, it is present in each measurement. In a SAS measurement, one determines the Fourier transform of the shape function of a structure that can be characterized by different parameters, and one of them is the radius of gyration. A coherent wave front 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 inversely proportional to the size of the structure, i.e. the larger the structure, the smaller the scattering angles and vice versa. The problem of using SAS for tomography is transforming 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 SAS, a broadening of the incident beam (wave) that is characteristic for the structure. The broadening  $B$  due to multiple scattering can be calculated as (Strobl et al. 2004a; Treimer et al. 2005d).

$$B = \sqrt{\int_{\text{path}} \frac{\sigma(x, y)N(x, y)}{R(x, y)^2} ds} \quad (114)$$

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

$$f_B(x, y) = \frac{\sigma(x, y)N(x, y)}{R(x, y)^2} ds \quad (115)$$

one can calculate the Radon transform as

$$P_\theta(t) = \int_{\text{path}} f_B(x, y(x, \theta, t)) ds \quad (116)$$

Thus, one can use position dependent broadening  $B$  as an imaging signal for tomography (Strobl et al. 2004a) (see also (Sabine and Bertram 1999)).

## Experimental Results: Refraction and Small Angle Scattering

As mentioned above, 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 the one into the left direction (This case is equivalent to the non-refracted one which also contributes to the 2D-refraction tomogram). One sees from (Eq. 99) that the individual refraction angles are of the order of  $10^{-5}$  rad, the sum  $\Delta$  is of course of the same order. To measure such small angles, a high angular resolution double crystal diffractometer (DCD) was used as shown in Fig. 50. The rocking curve of the analyser crystal of the DCD has a full width at half maximum (FWHM) of some sec of arc ( $\sim \mu\text{rad}$ ). Any small deviation (fractions of sec of arc) from this exact parallel position of monochromator and analyser crystals 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 in front of the analyser, which results in a decrease of the reflected intensity. This extreme angular sensitivity enables one to measure angles of the order of  $\mu$  rad and less with this instrument. To date, such DCD are used as ultra small angle instruments working in a so-called  $q$ -range ( $q$  = momentum transfer) of  $10^{-5} \text{ nm}^{-1} < q < 5 \cdot 10^{-3} \text{ nm}^{-1}$ , corresponding to particles sizes up to  $30 \mu\text{m}$ . However, one must distinguish absorption, refraction and small angle scattering to use them independently for imaging, despite the fact that they occur simultaneously.

To overcome these difficulties a special variant of the DCD can be taken, the bent analyser DCD. This instrument consists of a classical DCD of two perfect crystals, one as monochromator and the other as analyser. The monochromator can operate in the symmetric or asymmetric Bragg reflection, bent and un-bent mode. The analyser is a long perfect crystal, in which Bragg planes have an orientation in the crystal such that the (neutron) beam enters parallel to the crystal surface into the crystal (see Fig. 51).

Owing to a certain bending of the analyser crystal, the Bragg planes that are parallel to the monochromator ones will reflect neutrons, and the reflecting planes are reduced to a small volume. A position sensitive detector (512 channels, 300 channels can be used) behind the analyser now registers the position, i.e. where the neutrons are

reflected. Neglecting at first any absorption, rays, having a (slightly) different direction, will be Bragg reflected by another, different 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 ( $q$ ), whose width depends on the curvature of the crystal. The total angular or  $q$ -range is small (Darwin width,  $\sim 3.6 \cdot 10^{-4} \text{ nm}^{-1}$ ) if the analyser is not bent (radius is infinity). The rocking curve is spread over app. 300 channels, the  $q$ -range becomes large, if the analyser is strongly bent and the rocking curve becomes narrow, the intensity is distributed over a few (5–10) channels. The instrument is usually operated within a  $q$ -range of  $(5 \cdot 10^{-4} - 2.6 \cdot 10^{-2}) \text{ nm}^{-1}$ , which covers vice versa a wide angular range of ray deflections and thus particles sizes.

To use this DCD 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. However, at first one also has to correct the intensity distribution without a sample due to the different detection efficiency of all channels, the background and finally the absorption of the beam when it travels through the ( $Si$ ) analyser 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 ID-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 efficiencies between different channel numbers.

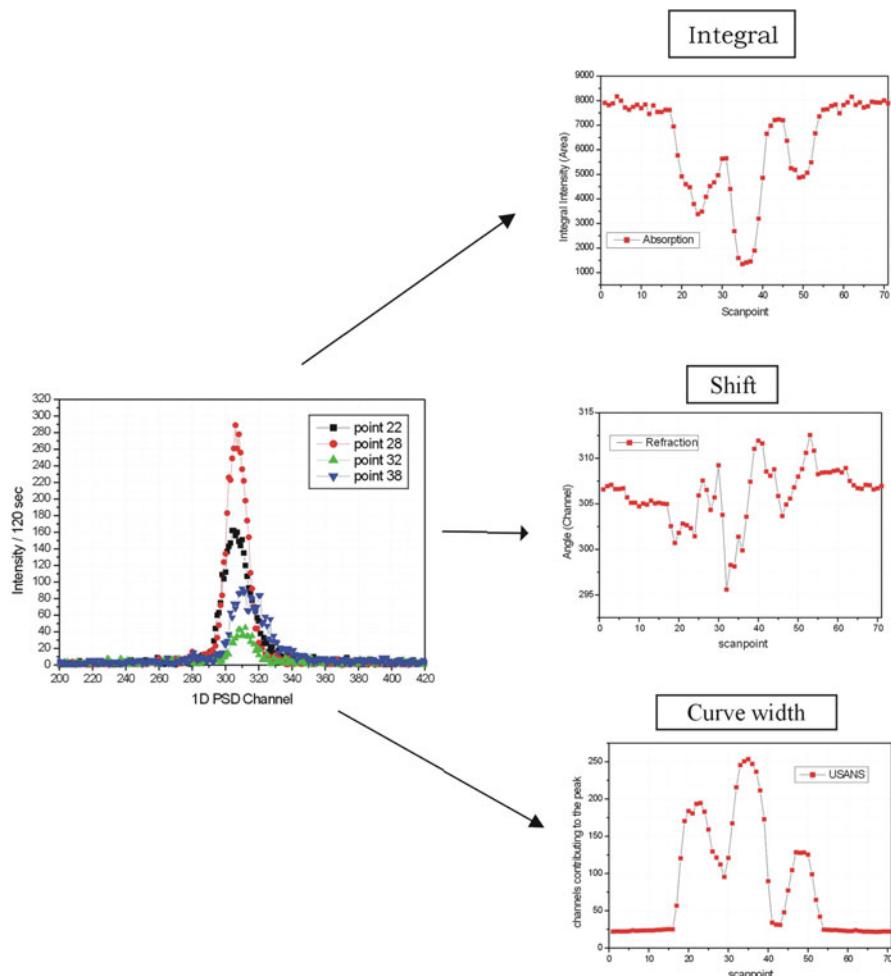
$$\frac{I_{\text{channel}}}{I_0} = \exp \left( - \int_{\text{length}=0}^{\text{channel}} \mu_{Si}(x) dx \right) \quad (117)$$

The absorption correction is then applied to the scattering curves and the data can be taken for further evaluations.

The sample is scanned step-wise translated perpendicular to the beam and the transmitted intensity is measured. This is done for a number of angular orientations of the sample with respect to incident radiation, usually 50–200 angles (The correct number of angles  $M$  is determined by number of scanning points  $N$ /projection,  $M = \pi/2N$ ; for monochromatic neutrons using a DCD,  $M$  must be a compromise between resolution and measuring time/projection) from  $0^\circ$  to  $180^\circ$ . 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  $I = \sum_{i=1}^{512} I_i$ , where only the 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  $\theta$  through the beam, one records projection points  $P_\theta(t)$ . To evaluate the rocking curves, one uses all projections  $P_\theta^{\text{absorb}}(t)$  and then the reconstruction formalism as described in section “[Mathematics of Computerized Tomography](#).”

Figure 56 shows for one orientation  $\theta$  four steps, i.e. rocking curves (RC) point 22, point 28, point 32 and point 38 from 70 steps (one projection has 70 steps). Each RC is evaluated with respect to its integrated intensity (‘Integral’), shift of the centre



**Fig. 56** Data recording and evaluation, one gets for each stepping point three numbers for three different interactions, absorption, refraction and small angle scattering and thus for one sample orientation, three projections and as a set (all stepping points) one gets the projections  $P_{\text{absorb}}(t)$ ,  $P_{\text{refrac}}(t)$ ,  $P_{\text{SANS}}(t)$

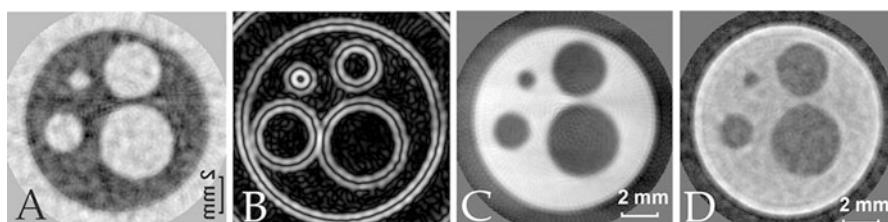
of the RC ('Shift') and width of the RC ('Curve width') yielding finally three projections, 'Integral', 'Shift' and 'Curve width'. The projection 'Integral' is simply the sum of all channels (corrected for background), to calculate the shift  $\Delta$  of each RC with respect to empty beam, the centre of gravity  $C_{\text{grav}}$  must be determined for each scanning point as

$$C_{\text{grav}} = \frac{\sum_{n=1}^{512} I_n \cdot n}{\sum_{i=1}^{512} I_n} = \frac{\sum_{n=1}^{512} I_n \cdot n}{I} \quad (118)$$

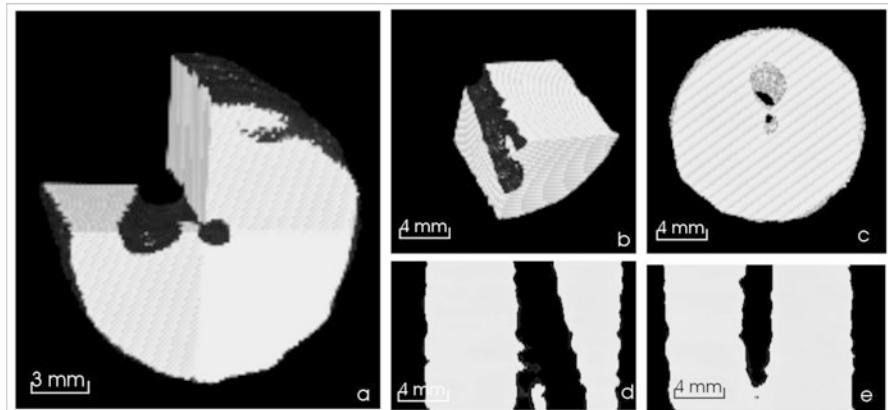
and the SANS data are derived from the broadening of the RC, which is determined by a certain number of channels, i.e. the number of channels that together contain 95% of the neutrons of the total rocking curve are plotted as a function of the scanning points. All these values yield projections  $P_{\text{Integral}}(t)$ ,  $P_{\text{Shift}}(t)$  and  $P_{\text{Curvewidth}}$ , which are used for image reconstruction.

Several samples have been tested, and one is given in Fig. 57. A cylindrical brass sample with holes of 1, 2, 3 and 4 mm could be reconstructed using absorption, refraction and US ANS data. Note the different contrast relating to the interaction. A 3D reconstruction could also be realized with this technique (Fig. 58).

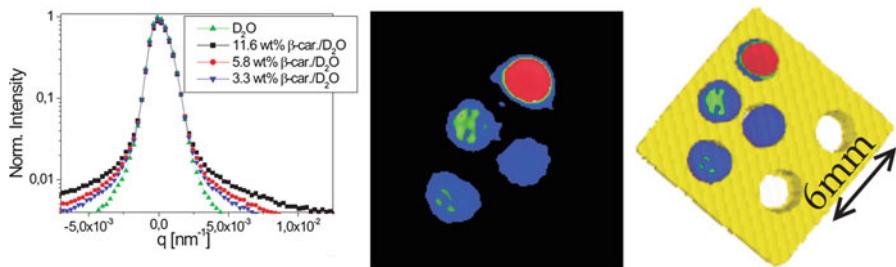
It was found that the broadening  $B$  (Eq. 114) of the small angle scattering pattern (rocking curve) can be taken as an imaging signal. However, the broadening must be defined in a unique manner that holds for all scanning points and orientations. For the un-effected rocking curve, the intensity is 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). The projection data for each scanning point were evaluated in the same way as the refraction data. The broadening is determined by the number of channels, left and right from the maximum, having together a certain integral intensity. This should be well above background; the integral intensity for SANS should be between 90% and 98%.



**Fig. 57** Different imaging signals, A = absorption, B = refraction, C = integrated B, D = USANS; see (Treimer et al. 2003)



**Fig. 58** 3D reconstruction from refraction data only (Treimer et al. 2005b; Strobl et al. 2004b)



**Fig. 59** Left graph: SANS curves for different concentrations of  $\beta$ -carotene, mean particle size  $\sim 150$  nm, middle image: 2D reconstruction from SANS data, right image: combined reconstruction using refraction and SANS data (Treimer et al. 2005b; Treimer et al 2005c)

Usually, a sample will give rise to all imaging signals, to absorption, refraction and small angle scattering. The big advantage of this technique is that all data can be taken from the same rocking curves from the same scans. There is no need to re-measure 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 shown in Figs. 58 and 59.

The special double crystal instrument is mainly used for small angle scattering experiments. As mentioned above, the lateral coherence length is in this case larger than the structures and 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  $\mu$ .

To investigate the sensitivity of this imaging signal, a small Al box with six holes was filled with different concentrations of  $\beta$ -carotene in  $D_2O$ . The small angle scattering of  $\beta$ -carotene is shown in Fig. 59 (the mean particle size  $\sim 150$  nm was

investigated before by means of SANS in order to know this value precisely). Combining small angle scattering reconstruction data with the refraction data, one obtains full information about the sample (different concentrations and exact shape).

## Neutron Grating Interferometry

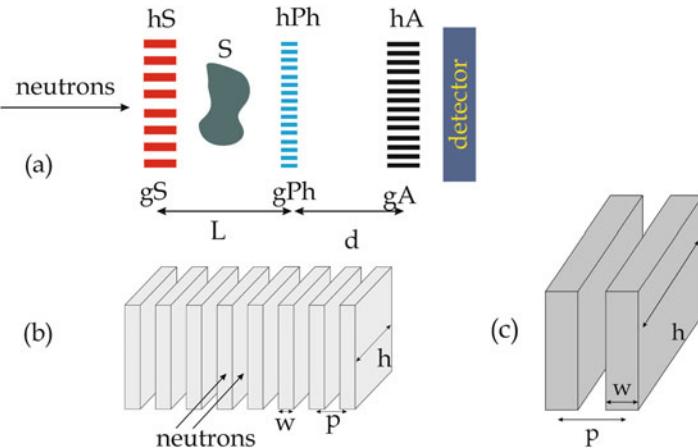
In 2006, F. Pfeiffer and co-workers presented an image technique based on X-ray grating interferometry, which was later applied to neutrons (Gruenzweig; Gruenzweig et al. 2008a, b; Strobl et al. 2008a; Pfeiffer et al. 2006; Reimann et al. 2016). In this section, a short overview is given and the first principles explained (for details see (Gruenzweig; Reimann et al. 2016)). The physical basis is given in 5.3. An object illuminated by a coherent wave front causes phase changes resulting in scattering, i.e. refraction, small angle scattering and/or diffraction, depending on the size of the object compared to the lateral coherence length. The small changes of beam direction by objects are called small angle neutron scattering and were first detected for neutron imaging with a double crystal diffractometer (Schaper 1996; Treimer et al. 2003). With this instrument the sample must be translated stepwise perpendicular to a pencil neutron beam (which defines the spatial resolution) and absorption, refraction and small angle scattering can be recorded simultaneously for one projection (see section “[Experimental Results: Refraction and Small Angle Scattering](#)”).

The use of a neutron grating interferometer (NGI) remarkably enhances the possibilities for imaging because it overcomes the lateral scanning. The sample does not have to be moved in the beam in order to gather data from the whole sample for a projection; however, some scanning must be performed to calculate transmitted intensity (TI), to get phase contrast, exactly spoken differential phase contrast (DPC) intensity and so-called dark-field contrast intensity (DFI). The principle of a grating interferometer is based on the Talbot effect. A plane wave passing through a phase grating creates an image of the grating itself at a certain distance downstream (Pfeiffer et al. 2006). The interferometer consists of three gratings, a source grating  $g_S$ , a phase grating  $g_{Ph}$  and an analysing grating  $g_A$ , see Fig. 60.

The source grating  $g_S$  creates a wave front in order to coherently illuminate the phase grating  $g_{Ph}$ , which had to be placed downstream far enough from  $g_S$  at a distance  $L$  (e.g.  $L \sim \text{meter}$ ), in order to get the necessary lateral coherence length (width) larger than the periodicity of phase grating  $g_{Ph}$ . The spacing (periodicity) for  $g_S$  is mm, whereas the spacings for  $g_{Ph}$  and  $g_A$  are of the order of  $\mu\text{m}$ . Then the necessary (lateral) coherence length can be estimated as

$$c_l = \lambda \cdot \frac{\phi}{\Delta\phi} \quad (119)$$

with  $\phi = L/D$  ratio. Here,  $D$  is (from  $L/D$ ) the spacing ( $p-w$ ) in (c) in Fig. 60, yielding  $(p-w) = l \text{ mm}$ ,  $\lambda = 0.4 \text{ nm}$  and  $L = 5 \text{ m}$ ,  $c_l = 20 \mu\text{m}$ . The gratings are defined by the period  $p = p_S$  (source grating),  $p = p_{Ph}$  (phase grating) and  $p = p_A$  (absorption grating) and by their heights  $h = h_S$  (source grating),  $p = h_{Ph}$  (phase grating) and



**Fig. 60** Layout of a neutron grating interferometer (NGI) using the Talbot effect.  $g_S$  is the source grating,  $g_{Ph}$  is the phase grating and  $g_A$  is the absorption grating, see text

$p = h_A$  (absorption grating). Each part of the source gratings  $g_S$  acts as a coherent source (width =  $w$ ), but is incoherent to its neighbouring bridges (see also (c) in Fig. 60). The phase shift of the phase grating  $g_{Ph}$  is given

$$\Delta\Phi = k \cdot (1 - n) \cdot h = -\lambda N b_c h \quad (120)$$

( $k = \frac{2\pi}{\lambda}$ ,  $n$  = index of refraction,  $N$  = atomic density of the grating material,  $b_c$  = coherent scattering length of the material (pure Silicon) and  $h$  = path length through the bridge (see also section “Phase-Based Imaging Signals”). The wave length dependent Talbot distances are  $d_{t,n_T}$

$$d_{t,n_T} = n_T \cdot k \cdot d_t, \quad d_t = \frac{2p_{Ph}^2}{\lambda} \quad (121)$$

Here,  $n_T$  is the Talbot order,  $k$  is the ratio of the smallest fractional Talbot distance and the total Talbot distance. Usually  $\Delta\Phi = \pi/2$  or  $\pi$  phase gratings are used. Downstream from the phase grating  $g_{Ph}$  the wave becomes phase modulated and by the Talbot effect transformed into intensity modulation in front of  $g_A$ .  $g_A$  scans of the intensity modulation in each detector pixel  $(x, y)$ , because the intensity oscillates in each detector pixel  $(x, y)$  due to a translation  $t_g$ ,  $g$  = any grating (Reimann et al. 2015, 2016; Reimann 2017)

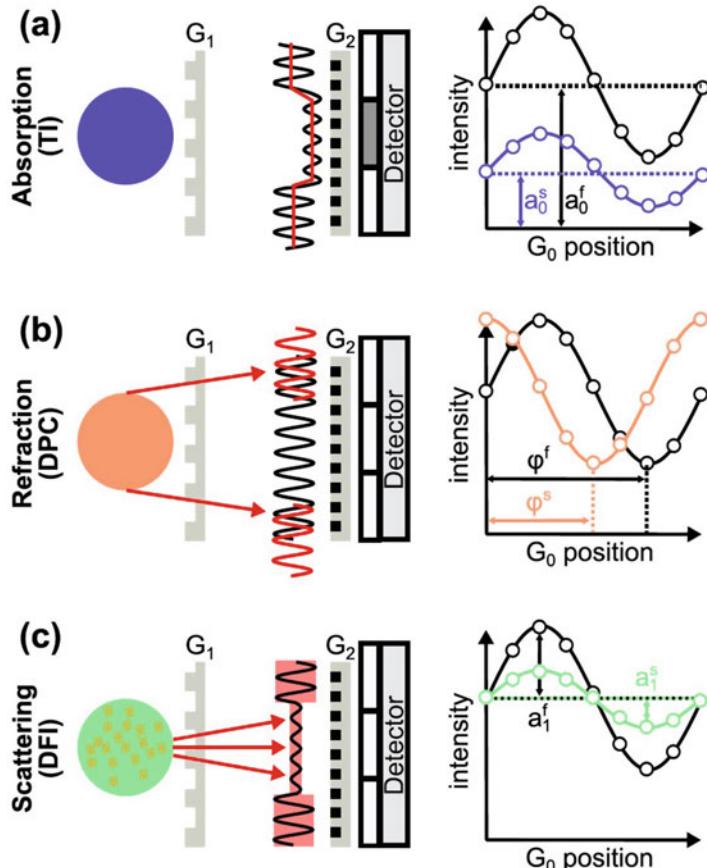
$$I(t_g, x, y) = a_0(x, y) + a_1(x, y) \cdot \cos \left( \frac{2\pi}{p_i} \cdot t_g - \varphi(x, y) \right) \quad (122)$$

with the offset  $a_0(x, y)$ , the amplitude  $a_1(x, y)$ , period width of grating  $p_i$  and phase  $\varphi$  of the oscillation. Data recording requires “empty beam images” for the parameters  $a_0(x, y)$ ,  $a_1(x, y)$  and  $\varphi(x, y)$  in order to calculate the absorption contrast from transmitted intensity

or transmitted image (TI), the refraction contrast or differential phase contrast image (DPC) and dark-field image contrast (DFI) as (Reimann et al. 2016)

$$\begin{aligned} \text{TI}(x, y) &= \frac{a_0^S(x, y)}{a_0^e(x, y)} \\ \text{DPC}(x, y) &= \varphi^S(x, y) - \varphi^e(x, y) \\ \text{DFI}(x, y) &= \frac{a_1^S(x, y)}{a_0^S(x, y)} \cdot \frac{a_0^e(x, y)}{a_1^e(x, y)} \end{aligned} \quad (123)$$

The superscript ‘e’ means empty beam image, i.e. without sample, the superscript ‘S’ means measurement with sample. As can be seen from (Eq. 123), the dark-field contrast DFI is based on the relative visibility, i.e. modulation amplitude and is found in the case of small angle scattering induced by the sample by sub-image-resolution structure sizes (Strobl et al. 2008a; Strobl 2014).



**Fig. 61** The nGI contrast mechanism (see Tommy Reimann, (Reimann et al. 2016; Reimann 2017))

Figure 61 shows (a) absorption, (b) refraction and (c) scattering by a sample causing different intensity oscillation  $I(t_g, x, y)$  if one grating is translated by  $t_g$  and hence generates contrast in the TI, DPC and DFI contrast channels, respectively (after Tommy Reimann, arXiv:1602.08846v1 [cond-mat.mtrl-sci] 29 Feb 2016) (Reimann et al. 2016; Reimann 2017). Comparing images in Fig. 57 (section “Experimental Results: Refraction and Small Angle Scattering”), one recognizes the corresponding TI, DPC and DFI images.

## Bragg Edge Imaging Using Energy Selective Neutron Technique

Residual stress is one of the main topics in the research field of nondestructive testing. It arises from manufacture processes. Heat treatments, melting welding, pressing, applying stress and strain forces on materials are origins of so-called residual stress. Often it is the starting point of flaws, such as fracture, cracks and other faults. It is the beginning of material failure and so contributes to a kind of ageing of matter. Therefore, it is important to know about the processes that expedite these processes driven by stress and strain forces. Strain  $\epsilon$  is commonly described as the ratio

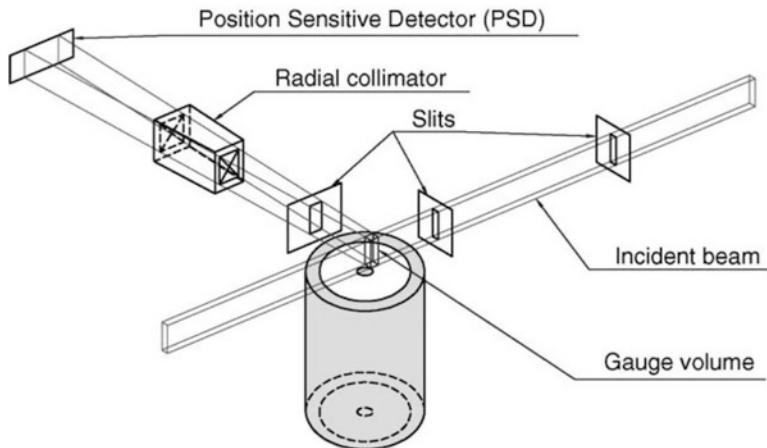
$$\epsilon = \frac{d_{\text{hkl}} - \bar{d}_{\text{hkl}}}{\bar{d}_{\text{hkl}}} = \frac{\Delta d_{\text{hkl}}}{\bar{d}_{\text{hkl}}} \quad (124)$$

where  $\bar{d}_{\text{hkl}}$  is the mean value of lattice spacing  $d_{\text{hkl}}$  in a given volume of a sample. The idea is to use changes of lattice spacing in order to detect residual stress in a sample. Residual stress itself can arise from different defects (misfits) and has many origins and length scales (details see in (Withers et al. 2008)). In this section, energy selective neutron imaging is described, which utilizes Bragg-edges of crystalline materials in order to obtain microstructural information related to the lattice spacing, such as phase, texture and strain.

In (Santisteban et al. 2001) it was demonstrated, that by means of neutron time of flight (TOF) transmission experiments, Bragg edges could be determined with a precision of  $\frac{\Delta d_{\text{hkl}}}{d_{\text{hkl}}} \sim 10^{-5}$ , which was applied to determine shape, magnitude and location of the Bragg edges in the transmission spectrum of polycrystalline materials. The possibility of getting information about the stress state, texture and phases present in the material was pointed out by (Santisteban et al. 2002).

## Theory

Matter consists of atoms that form small crystals in solid state; thus, polycrystalline materials. Strain measurements use the fact that all small crystals have a constant mean inter planar spacing  $d$  that can be precisely determined with Bragg's law  $2d \cdot \sin(\theta) = \lambda$ . (see Santisteban et al. 2002). This can be done with X-rays and neutrons, where the volume of interest (gauge volume) is rather small ( $1\text{mm}^3$ ) using classical methods of diffraction.



**Fig. 62** The gauge volume of sampling or gauge volume is defined by slits in the incident and diffracted beams to be about  $1 \text{ mm} \times 1 \text{ mm} \times 8 \text{ mm}$  (sketch from C.C. Aydiner et al. (2005))

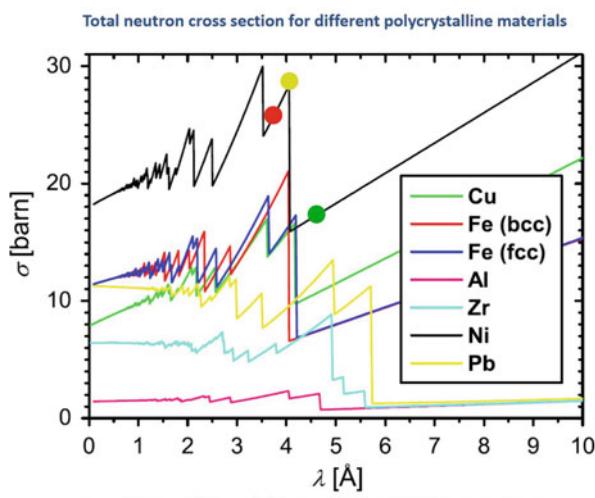
Diffraction and time-of-flight are standard methods, but for nondestructive testing they are quite time consuming, as can be estimated from Fig. 62. However, owing to developments in neutron imaging concerning imaging signals other than absorption, such as (all kinds of) phase-based and polarized neutron imaging, and also driven by development of improved or new detector systems allowing spatial resolution of less than  $50/\text{nm}$ , neutron imaging is also applied to stress and strain measurements. In X-ray and neutron scattering, lattice spacing is measured by Bragg diffraction of several lattice planes ( $hkl$ ) with a monochromatic beam, the corresponding Bragg angle  $\theta$  is defined by

$$n\lambda = 2d_{hkl} \sin(\theta) \quad (125)$$

From Bragg's law, the Bragg angle is limited to  $\Theta = 90^\circ$ , i.e. for a given wave length  $\lambda$ , ( $n = 1$ , (Eq. 125)) and for  $\theta = 90^\circ$ ,  $\lambda = 2d_{hkl}$ . Using a full neutron spectrum with  $(0.1 \text{ nm} \leq \lambda \leq 1 \text{ nm})$ , one observes so-called Bragg edges for the transmitted beam through a sample (see Fig. 63). Owing to the wave length spectrum (polychromatic neutrons), neutrons of different wavelengths can be Bragg reflected by a polycrystalline sample leading to characteristic transmission spectra through polycrystalline samples with wave length dependent Bragg edges (Fig. 63). Neutrons which do not fulfil the Bragg equation can pass through the sample without scattering. The position of these Bragg edges is directly related to the corresponding d-spacing of the lattice planes of the polycrystalline material. The problem for such experiments is the correct determination of the position of a Bragg edge which depends on the edge height  $h$  (and uncertainty  $\Delta h$ ), on wavelength,  $\Delta\lambda/\lambda$ , available intensity, background, etc.

The measurements are simple in performance; however, the data evaluation is not. As an example, the Austenite and  $\alpha$ -Martensite phases of iron are given below.

**Fig. 63** Bragg edges for different crystals after (Josic et al. 2011)



Josic, L., Lehmann, E., & Kaestner, a. (2011). Nucl. Instruments Methods Phys. Res. 651, 166–170.

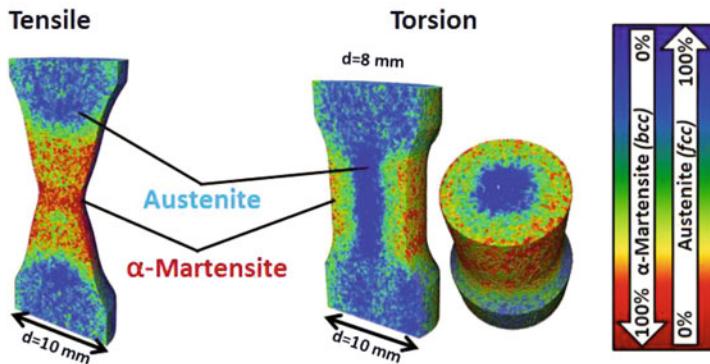
## Application

The following experiment is based on the papers (Woracek et al. 2011; Woracek 2015), which clearly demonstrate the potential of neutron imaging in nondestructive testing (see Fig. 64). They present experiments of nondestructive three-dimensional (3D) mapping of crystallographic phase distributions of Austenite and  $\alpha$ -Martensite phases of iron. The spatial resolution is of the order of  $50 \mu\text{m}$ , which is much better than can be measured with the classical diffraction method.

The sample is placed in a collimated monochromatic neutron beam where the incident wave length can be varied from  $\lambda_{\min}(0.2 \text{ nm})$  to  $\lambda_{\max}(0.6 \text{ nm})$ . The collimation ( $L/D$  ratio) mainly determines the spatial resolution of the reconstruction. The sample is as a whole illuminated, so that from each volume element of the sample, a Bragg edge spectrum can be evaluated. The Bragg edge spectrum from the Austenite phase of iron differs from the  $\alpha$ -Martensite phase so both phases present in the sample could be localized. For tomography, the sample is rotated  $360^\circ$  at two wave lengths, one before ( $0.41 \text{ nm}$ ) and after ( $0.43 \text{ nm}$ ) the Bragg cut-off corresponding to the austenitic phase yielding a remarkable contrast between the Austenite phase and  $\alpha$ -Martensite phase. The reconstruction was performed for the parallel beam option using filtered back projection (FBP), and the result is shown in Fig. 64.

One can clearly distinguish both phases for tensile force and torsion force. The necking region of the tensile sample is fully transformed to Martensite, whereas it stays in the Austenite phase in the case of torsion. Thus, the radial dependence of the phase transformation in the torsion sample is clearly depicted and quantified.

Determining bulk residual elastic strains is a topic for predictive modelling and materials testing. There are numerous approaches to obtain this information, including those concerning bulk residual stress described by (Aydiner et al. 2005; Josic



**Fig. 64** 3D reconstruction of the phase fractions for the tensile and torsion sample see (Woracek et al. 2015)

et al. 2011), and a different method is described in (Kirkwood et al. 2015). Because the measuring process with classical methods is very time-consuming and frequently involves destructive sectioning of the sample, neutron imaging is certainly a powerful alternative for nondestructive testing and engineering.

## References

- Andersen AH, Kak AC (1984) Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrason Imaging* 6:81–94
- Anderson IS, et al (eds) (2009) Neutron imaging and applications, neutron scattering applications and techniques. Springer Science and Business Media, LLC. [https://doi.org/10.1007/978-0-387-78693-3\\_6](https://doi.org/10.1007/978-0-387-78693-3_6)
- Annet JF (2013) Superconductivity, superfluids and condensates. Oxford University Press, Chapter 3.7
- ASTM Standards designation: E 803-91 (Reapproved 1996) Standard test method for determining the L/D ratio of neutron radiography beams. <https://www.astm.org/Standards/E803.htm>
- Aull S, Ebrahimi O, Karakas N, Knobloch J, Kugeler O, Treimer W (2012) Suppressed Meissner-effect in niobium: visualized with polarized neutron radiography. *J Phys Conf Ser* 340 (012001):1–7. <https://doi.org/10.1088/1742-6596/340/1/012001>
- Aydiner CC, Uetuendag E, Clausen B, Hanan JC, Winholtz RA, Bourke MAM, Peker A (2005) Residual stresses in a bulk metallic glass – stainless steel composite. *Mater Sci Eng A* 399:107–113
- Bacon GE (1975) Neutron diffraction., Chapter 3, 3rd edn. Clarendon Press, Oxford
- Badurek G (2011) NESY Winter School, <http://planner2011.unileoben.ac.at/fileadmin/shares/planner2011/docs/private/29-Badurek.pdf>
- Badurek G, Hochhold M, Leeb H (1997a) Neutron magnetic tomography – a novel technique. *Physica B* 234–236:1171–1173
- Badurek G, Hochhold M, Leeb H, Buchelt R, Korinek F (1997b) A proposal to visualize magnetic domains within bulk materials. *Physica B* 241–243:1207–1209
- Badurek G, Hochhold M, Leeb H, Buchelt R, Korinek F (1998) A proposal to visualize magnetic domains within bulk materials. *Physica B* 241–243:1207–1209

- Badurek G, Buchelt RJ, Kroupa G, Baron M, Villa M (2000) Permanent magnetic field-prism polarizer for perfect crystal neutron interferometers. *Physica B* 283:389–392
- Boeni P, Miinzer W, Ostermann A (2009) *Phys B Condens Matter* 404(17):2620–2623
- Bracewell RN (1956) *Aust J Phys* 9:198
- Brenizer JS (2013) A review of significant advances in neutron imaging from conception to the present. *Phys Procedia* 43:10–20
- Chadwick J (1932) Possible existence of a neutron. *Nature* 129:312
- Deans SR (1983) The radon transform and some of its applications. Wiley, New York. ISBN 13: 978-486-46421-7 and 10: 0-486-46421-2
- Dhiman I, Ziesche R, Anand VK, Riik L, Song G, Islam ATMN, Tanaka I, Treimer W (2017) Thermodynamics of Meissner effect and flux pinning behavior in the bulk of single-crystal  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$  ( $x=0.09$ ). *Phys Rev B* 96:104517-1–104517-10
- Ernst C (1997) Untersuchung von  $\pi$  – Flipfern mit kalten Neutronen. Diploma thesis, supervisor W. Treimer, FB II, Technische Fachhochschule Berlin
- Glatter O, Kratky O (1982) Small angle X-ray scattering. Academic Press, London
- Gruenzweig C (2009) Neutron grating interferometry for imaging magnetic structures in bulk ferromagnetic materials. Doctorate thesis ETH Zuerich, Swiss, Dissertation ETH Nr. 18612
- Gruenzweig C, David C, Bunk O, Dierolf M, Frei G, Kähnhe G, Schäffer R, Pofahl S, Räynnnow HMR, Pfeiffer F (2008a) Bulk magnetic domain structures visualized by neutron dark-field imaging. *Appl Phys Lett* 93:112504
- Gruenzweig C, David C, Bunk O, Dierolf M, Frei G, Kuehne G, Kohlbrecher J, Schaefer R, Lejcek P, Räynnnow HMR, Pfeiffer F (2008b) *Phys Rev Lett* 101:025504
- Halpern O, Johnson MH (1939) On the magnetic scattering of neutrons. *Phys Rev* 55:898
- Hansen PC and Saxild-Hansen M (2010). AIR Tools - A MATLAB Package of Algebraic Iterative Reconstruction Techniques: Version 1.0 for Matlab 7.8. Kgs. Lyngby, Denmark: Technical University of Denmark, DTU Informatics, Building 321. IMM-Technical Report-2010-15
- Hansen PC, Saxild-Hansen M (2012) AIR tools – a MATLAB package of algebraic iterative reconstruction methods. *J Comput Appl Math* 236:2167–2178
- Hayter JB, Mook HA (1989) Discrete thin-film multilayer design for X-ray and neutron supermirrors. *J Appl Crystallogr* 22:35
- Herman GT (1980) Image reconstruction from projections in computer science and applied mathematics. Academic Press, New York
- Herzig C (1997) Experimentelle Realisierung von 3D-tomographien aus Neutronenradiographien. Diploma thesis, supervisor W. Treimer, FB II, Technische Fachhochschule Berlin
- Hilger A (2010) Charakterisierung magnetischer Strukturen durch bildgebende Verfahren mit kalten Neutronen. Doctoral thesis, D 83, Technical University Berlin
- Hochhold M, Leeb H, Badurek G (1996) Tensorial neutron tomography: a first approach. *J Magn Magn Mater* 157–158:575
- Iavarone M, Moore S, Fedor J, Ciocys S, Karapetrov G, Pearson J, Novosad V, Bader S (2014) Visualizing domain wall and reverse domain superconductivity. *Nat Commun* 5:4766
- Jericha E, Szeywerth R, Leeb H, Badurek G (2007) Reconstruction techniques for tensorial neutron tomography. *Physica B* 397:159–161
- Josic L, Lehmann E, Kaestner A (2011) Energy selective neutron imaging in solid state materials science. *Nucl Instrum Methods Phys Res Sect A* 651:166–170
- Kaczmarz S (1937) Angenäherte Aufloesung von Systemen linearer Gleichungen. *Bulletin de l'Académie Polonaise des Sciences et Lettres* A35:355–357
- Kak AC, Slaney M (1999) Principles of computerized tomographic imaging. IEEE Press, Inc, New York
- Kardjilov N, Manke I, Strobl M, Hilger A, Treimer W, Meissner M, Krist T, Banhart J (2008) *Nat Phys* 4:399–403
- Keimer B, Belk N, Birgeneau RJ, Cassanho A, Chen CY, Greven M, Kastner MA, Aharoni A, Endoh Y, Erwin RW, Shirane G (1992) Magnetic excitations in pure, lightly doped, and weakly metallic  $\text{La}_2\text{CuO}_4$ . *Phys Rev B* 46:14034
- Kirkwood HJ, Zhang SY, Tremsin AS, Korsunsky AM, Baimpas N, Abbey B (2015) Neutron strain tomography using the radon transform. *Mater Today Proc* 2S:S414–S423

- Kraft J (1996) Konstruktion, Aufbau und Test einer Neutronenpolarisations- und analyseeinrichtung. Diploma thesis, supervisor W. Treimer, FB II, Technische Fachhochschule Berlin
- Krist T, Kennedy SJ, Hick TJ, Mezei F (1998) Physica B 241–243:82–85
- Leeb H, Hochhold AM, Badurek G, Buchelt RJ, Schricker A (1998) Neutron magnetic tomography: a feasibility study. Aust J Phys 51:401–413
- Leeb H, Szeywerth R, Jericha E, Badurek G (2005) Towards manageable magnetic field retrieval in bulk materials. Physica B 356:187–191
- Maass P, Treimer W, Feye-Treimer U (1992) Tomographic Methods for 2D Reconstructions with the Double Crystal Diffractometer. *IMPACT*, 4(3):250–268
- Magnetic form factors: see, <https://www.ill.eu/sites/csl/ffacts/ffachtml.html>
- Manke I, Kardjilov N, Hilger A, Strobl M, Dawson M, Banhart J (2009) Polarized neutron imaging at the CONRAD instrument at Helmholtz Centre Berlin. Nucl Instrum Methods Phys Res Sect A 605(1–2):26–29
- Manke I, Kardjilov N, Schaefer R, Hilger A, Strobl M, Dawson M, GrAijnzweig C, Behr G, Hentschel M, David C et al (2010) Three-dimensional imaging of magnetic domains. Nat Commun 1:125
- Messiah A (1999) Quantum mechanics. Dover Publishing, Inc, Mineola., Chapter II
- Mezei F (1972) Neutron spin echo: a new concept in polarized thermal neutron techniques. Z Physik 255(2):146–160
- Mezei F (1976) Novel polarized neutron devices: supermirror and spin component amplifier. Commun Phys 1:81
- Müller KA, Bednorz JG (1987) The discovery of a class of high-temperature superconductors. Science 237:1133
- NIST Neutron scattering length and cross sections. <https://www.ncnr.nist.gov/resources/n-lengths/>, <https://physics.nist.gov/cuu/Constants/index.html>
- NIST Center for Neutron Research. <https://ncnr.nist.gov/resources/n-lengths/elements/fe.html>
- Pfeiffer F, Weitkamp T, Bunk O, David C (2006) Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources. Nat Phys 2:258
- Poole CP, Farach HA, Prozorov R (2007) Superconductivity, 2nd edn. Elsevier Ltd, p 53. Copyright 1995–2007
- Radon J (1917) Ueber die Bestimmung von Funktionen durch ihre Integralwerte lAd'ngs gewisser Mannig-faltigkeiten. Berichte Saechsischer Akademie Der Wissenschaften 29:262
- Rauch H, Werner SA (2000) Neutron interferometry in lessons in experimental quantum mechanics. Clarendon Press, Oxford, p 3
- Reimann T (2017) Vortex matter beyond SANS, Neutron studies of vortex structures covering a length scale of 0.01 to 10 μm. Doctorate thesis, Technische Universitaet Muenchen Physik-Department, Institut E21, Forschungs-Neutronenquelle Heinz Maier-Leibnitz (FRM II)
- Reimann T, Muehlbauer S, Schulz M, Betz B, Kaestner A, Pipich V, Boeni P, Gruenzweig C (2015) Visualizing the morphology of vortex lattice domains in a bulk type-II superconductor. Nat Commun 6:8813
- Reimann T, Muehlbauer S, Horisberger M, Boeni P, Schulz M (2016) The new neutron grating interferometer at the ANTARES beamline: design, principles, and applications. J Appl Crystallogr 49:1488–1500
- Rekveldt MT (1973) Study of ferromagnetic bulk domains by neutron depolarisation in three dimensions. Z Physik 259:391–410
- Rekveldt MT (1976) Correlation in ferromagnetic domain structures studied by means of the neutron depolarization technique. J Magn Magn Mater 1:342
- Rosenfeld A, Kak AC (1982) Digital picture processing, computer science and applied mathematics. Academic Press Inc., New York
- Sabine TM, Bertram WK (1999) The use of multiple-scattering data to enhance small-angle neutron scattering experiments. Acta Crystallogr A 55:500
- Sales M, Strobl M, Shinohara T, Tremsin A, Theil Kuhn L, Lionheart WRB, Desai NM, Dahl AB, Schmidt S (2018) Three dimensional polarimetric neutron tomography of magnetic fields. Sci Rep 8:2214. <https://doi.org/10.1038/s41598-018-20461-7>
- Santisteban JR, Edwards L, Steuwer A, Withers PJ (2001) Time-of-flight neutron transmission diffraction. J Appl Crystallogr 34:289–297

- Santisteban JR, Edwards L, Fitzpatrick ME, Steuwer A, Withers PJ (2002) Engineering applications of Bragg-edge neutron transmission. *Appl Phys A Mater Sci Process* 74(Suppl):S1433–S1436. <https://doi.org/10.1007/s003390101241>
- Schaerpel O (1989) Comparison of theoretical and experimental behaviour of supermirrors and discussion of limitations. *Physica B: Condensed Matter*, Volumes 156–157, January–February, 631:638
- Schaerpel O. The spin of the neutron as a measuring probe.
- Schaper J (1996) Untersuchungen zum Refraktionskontrast bei Tomographien mit themischen Neutronen. Diploma thesis, supervisor W. Treimer, FB II, Technische Fachhochschule Berlin
- Schillingen B (1996) 3D computer tomography with thermal neutrons at FRM garching. *J Neutron Res* 4:57–63
- Schillinger B, Gebhard R, Haas B, Ludwig W, Rausch C, Wagner U (1996) 3D neutron tomography in material testing and archaeology. In: Proceedings of the 5th world conference on neutron radiography, Berlin, printed 1997, pp 688–693
- Schulz M (2010) Radiography with polarised neutrons. Doctorate thesis TU Muenchen, Physik Department E21 (Lehrstuhl fuer Experimentalphysik III)
- Schulz M, Neubauer A, Masalovich S, Muehlbauer M, Calzada E, Schillinger B, Pfeiderer C, Boeni P (2010a) Towards a tomographic reconstruction of neutron depolarization data. *J Phys Conf Ser* 211:012025
- Schulz M, Neubauer A, Muehlbauer M, Calzada E, Schillinger B, Pfeiderer C, Boeni P (2010b) Polarized neutron radiography with a periscope. *J Phys Conf Ser* 200:112009
- Sears VF (1992) Neutron scattering lengths and cross sections. *Neutron News* 3(3):26–37. <https://doi.org/10.1080/10448639208218770>
- Shi X, Lin PV, Sasagawa T, Dobrosavljević V, Popović D (2014) Two-stage magnetic-field-tuned superconductor insulator transition in underdoped  $La_{2-x}Sr_xCuO_4$ . *Nat Phys* 10:437
- Shinohara T, Sakai K, Ohi M, Kai T, Harada M, Oikawa K, Maekawa F, Suzuki J, Oku T, Takata S, Aizawa K, Arai M, Kiyanagi Y (2011) Quantitative magnetic field imaging by polarized pulsed neutrons at J-PARC. *Nucl Instrum Methods Phys Res A* 651:121–125
- Shull CG, Wollan EO, Koehler WC (1951) Neutron scattering and polarization by ferromagnetic materials. *Phys Rev* 84(5):912–921
- Strobl M (2014) General solution for quantitative dark-field contrast imaging with grating interferometers. *Sci Rep* 4:7243
- Strobl M, Treimer W, Hilger A (2004a) Small angle scattering signals for (neutron) computerized tomography. *Appl Phys Lett* 85:448
- Strobl M, Treimer W, Hilger A (2004b) First realisation of a three-dimensional refraction contrast computerised neutron tomography. *Nucl Instrum Methods Phys Res B* 222:653
- Strobl M, Treimer W, Walter P, Keil S, Manke I (2007) Magnetic field induced differential neutron phase contrast imaging. *Appl Phys Lett* 91:254104
- Strobl M, Grünzweig C, Hilger A, Manke I, Kardjilov N, David C, Pfeiffer F (2008a) Neutron dark-field tomography. *Phys Rev Lett* 101:123902
- Surkau R, Becker J, Ebert M, Grossmann T, Heil W, Hofmann D, Humblot H, Leduc M, Otten EW, Rohe D, Siemensmeyer K, Steiner M, Tasset F, Trautmann N (1997) Realization of a broad band neutron spin filter with compressed, polarized  $^3\text{He}$  gas. *Nucl Instrum Methods Phys Res A* 384 (2–3):444–450
- Takagi H, Ido T, Ishibashi S, Uota M, Uchida S, Tokura Y (1989) Superconductor-to-non-superconductor transition in  $(La_{1-x}Sr_x)_2CuO_4$  as investigated by transport and magnetic measurements. *Phys Rev B* 40:2254
- Tarascon J-M, Greene L, McKinnon W, Hull G, Geballe T (1987) Superconductivity at 40 K in the oxygen-defect perovskites  $La_{2-x}Sr_xCuO_{4-y}$ . *Science* 235:1373
- Treimer W (1998) On double crystal diffractometry. *Cryst Res Technol* 33:643
- Treimer W (2014) Radiography and tomography with polarized neutrons. *J Magn Magn Mater* 350:188–198
- Treimer W, Feye-Treimer U (1998) Two-dimensional reconstruction of small angle scattering pattern from rocking curves. *Physica B* 241–243:128–1230
- Treimer W, Feye-Treimer U (2011) On coherence in neutron imaging. *Nucl Instrum Methods Phys Res A* 651:117–120

- Treimer W, Maass P, Strothmann H, Feye-Treimer U (1991) High-resolution neutron small-angle scattering with a double-crystal diffractometer and 2D reconstruction. *Physica B* 174:532–536
- Treimer W, Höfer A, Strothmann H (1997) The use of a multi-double-crystal diffractometer to investigate nickel domains. *J Appl Crystallogr* 30:849–853
- Treimer W, Feye-Treimer U, Herzig C (1998) On neutron tomography. *Physica B* 241–243:1297–1203
- Treimer W, Strobl M, Hilger A, Seifert C, Feye-Treimer U (2003) Refraction as imaging signal for computerized (neutron) tomography. *Appl Phys Lett* 83:389
- Treimer W, Hilger A, Kardjilov N, Strobl M (2005a) Review about old and new imaging signals for neutron computerized tomography. *Nucl Instrum Methods Phys Res Sect A* 542(2005):367–375
- Treimer W, Kardjilov N, Feye-Treimer U, Hilger A, Manke I, Strobl M (2005b) Absorption- and phase-based imaging signals for neutron tomography. In: Kramer B (ed) *Advances in solid state physics*, vol 45. Springer Verlag, Berlin/Heidelberg/New York, pp 407–420. ISSN 1438-4329
- Treimer W, Kardjilov N, Feye-Treimer U, Hilger A, Manke I, Strobl M (2005c) *Adv Solid State Phys* 45. Ed. Bernhard Kramer, Springer Verlag, pp 407–420
- Treimer W, Strobl M, Hilger A, Peschke HJ (2005d) Neutron tomography using small angle scattering data. *IEEE* 52(1):386–388
- Treimer W, Seidel SO, Ebrahimi O (2010) Neutron tomography using a crystal monochromator. *Nucl Instrum Methods Phys Res A* 621:502–505
- Treimer W, Ebrahimi O, Karakas N, Seidel SO (2011) PONTO- an instrument for imaging with polarized neutrons. *Nucl Instrum Methods Phys Res A* 651:53–56. <https://doi.org/10.1016/j.nima.2011.01.009>
- Treimer W, Ebrahimi O, Karakas N (2012a) Observation of partial Meissner effect and flux pinning in superconducting lead containing non-superconducting parts. *Appl Phys Lett* 101:162603-1–162603-4
- Treimer W, Ebrahimi O, Karakas N, Prozorov R (2012b) Polarized neutron imaging and three-dimensional calculation of magnetic flux trapping in bulk of superconductors. *Phys Rev B* 85 (18):184522-1–184522-9
- Treimer W, Ebrahimi O, Karakas N (2014) PONTO: an instrument for high resolution radiography and tomography with polarized neutrons. *Neutron News* 25(2):15–18. <https://doi.org/10.1080/10448632.2014.902698>
- Tremsin AS, McPhate JB, Vallerga JV, Siegmund OHW, Feller WB, Lehmann E, Butler LG, Dawson M (2011) High-resolution neutron microtomography with noiseless neutron counting detector. *Nucl Instrum Methods Phys Res A* 652:400–403
- Van Hemelryck Tessa, Wuyts Sarah, Goossens Maggie, Batenburg Kees, Joost, Sijbers Jan (2007) The implementation of iterative reconstruction algorithms in Matlab. Masters thesis, University College of Antwerp
- Withers PJ, Turski M, Edwards L, Bouchard PJ, Buttle DJ (2008) Recent advances in residual stress measurement. *Int J Press Vessel Pip* 85:118–127
- Woracek R (2015) Energy selective neutron imaging for the characterization of polycrystalline materials. PhD dissertation, University of Tennessee. [http://trace.tennessee.edu/utk\\_greddiss/3375](http://trace.tennessee.edu/utk_greddiss/3375)
- Woracek R, Penumadu D, Kardjilov N, Hilger A, Strobl M, Wimpory RC, Manke I, Banhart J (2011) Neutron Bragg-edge-imaging for strain mapping under in situ tensile loading. *J Appl Phys* 109:093506
- Woracek R, Penumadu D, Kardjilov N, Hilger A, Boin M, Banhart J, Manke I (2015) Neutron Bragg edge tomography for phase mapping. *Phys Procedia* 69:227–236



# Positron Annihilation

35

Luca Chiari and Masanori Fujinami

## Contents

Introduction .....	1302
Role and Formation of Defects in a Lattice Crystal .....	1303
Background of Positron Annihilation Spectroscopy .....	1304
Brief History of Positron Annihilation .....	1304
Principles of the Detection of Defects .....	1305
Trapping Model .....	1308
Positronium Formation in Molecular Systems .....	1311
Experimental Techniques .....	1313
Positron Sources .....	1313
Positron Annihilation Lifetime Spectroscopy .....	1314
Doppler Broadening Spectroscopy .....	1318
Variable Energy Positron Annihilation Spectroscopy .....	1323
Application Examples .....	1327
Metals .....	1327
Semiconductors .....	1333
Polymers .....	1336
Conclusions and Future Perspectives .....	1341
References .....	1342

### Abstract

Positron annihilation spectroscopy is a nondestructive technique that has been extensively applied in recent decades to detect the presence of vacancy-type defects in a large variety of materials. It is particularly suitable to investigate the size and concentration of vacancy-type defects at various depths in metals, alloys, semiconductors, porous materials, and polymers. In this chapter, the main

---

L. Chiari · M. Fujinami (✉)

Department of Applied Chemistry and Biotechnology, Graduate School of Engineering,  
Chiba University, Chiba, Japan

e-mail: [luca.chiari@chiba-u.jp](mailto:luca.chiari@chiba-u.jp); [fujinami@faculty.chiba-u.jp](mailto:fujinami@faculty.chiba-u.jp)

experimental techniques that take advantage of positron annihilation are reviewed, the data analysis procedures are discussed, and the information obtained in this kind of measurements is described. Typical applications of these methods are illustrated through examples of investigations on various kinds of materials. Advantages, present limitations, and potential future developments of these techniques are discussed in detail.

---

## Introduction

Most of the visible universe is made of matter, namely, atoms and molecules composed of particles like protons, neutrons, and electrons. However, according to the laws of physics, one can also define a material composed of antimatter to the corresponding particles of ordinary matter: that is antimatter. The positron is the antiparticle of the electron: they share the same mass but possess opposite electric charge. The positron is interesting not just from a fundamental point of view, for example, to study the forces governing its interaction with matter, but it possesses numerous other applications in astrophysics, solar physics, particle physics, biomedicine, and last, but not least, materials science.

Precisely in materials science, positrons are at the heart of so-called positron annihilation spectroscopy (PAS), a range of techniques that has been widely used to study vacancy-type defects in many kinds of materials since the late 1960s. The aim of this chapter is to introduce the basic concepts behind the experimental methods of positron annihilation and briefly give some examples of their applications to identify and improve our understanding of the role of defects in the physical, optical, and electrical properties of various types of materials.

This chapter is organized as follows. First, an introduction to defects in solid crystals is given. Then, the history and methods of positron annihilation are briefly overviewed. After introducing the behavior of positrons in solids and shortly overviewing the theoretical principles behind the dynamics, the details of the key PAS methods are described, namely, positron annihilation lifetime spectroscopy (PALS), and Doppler broadening spectroscopy (DBS). Further, variable energy positron annihilation spectroscopy (VEPAS) is mentioned for the analysis of surface layers (with a thickness of up to  $\sim 1 \mu\text{m}$ ). The strengths and weaknesses of the experimental methods and analysis techniques are discussed in order to provide a physical interpretation of the positron results. Some examples of applications of the experimental PAS methods to the investigation of various materials and the detection of defects are also presented. The aim of these examples is to show some cases of how the PAS techniques work in practice and how information about defects can be obtained from this kind of measurements. Finally, the present challenges and possible future directions in research with positrons are discussed.

It is important to note that this chapter does not aim to comprehensively describe all PAS techniques and the related positron work on defects in solids. It rather gives an overview of the main positron methods and a selection of the work published in the field. To cover most of the published works on positron annihilation in solids, we

refer the interested reader to earlier books (e.g., Charlton and Humberston 2001; Coleman 2000; Hautojärvi 1979; Jean et al. 2003; Krause-Rehberg and Leipner 1999; Mallon and Schrader 2003; Mogensen 1995) and reviews (e.g., Gidley et al. 2006; Jean et al. 2013; Schultz and Lynn 1988; Siegel 1980; Singh 2016; Tuomisto and Makkonen 2013) and references therein.

## Role and Formation of Defects in a Lattice Crystal

Atoms in real solids are not arranged according to a perfect crystalline pattern but rather contain imperfections to some extent. Defects in crystalline solids are static interruptions in the regular pattern of the lattice crystal. They may arise because of inherent limitations in the material production method, when the material is subject to mechanical stress or by exposure to radiation, etc. Given that atoms are relatively immobile in solids, once imperfections have been introduced into the crystal, for instance, during its formation, processing, or use, they are hard to remove. The presence of defects is very important because they can strongly affect the properties of materials. In particular, it is known that materials are often stronger when they contain defects. In fact, precisely because real materials are not perfect crystals, it is possible to tailor their properties for a large variety of applications. The development of a diversified range of materials is critical in order to meet the needs of emerging technologies and the growing demand in the materials engineering industry.

Crystallographic defects exist in a large variety of sizes and kinds. They can be primarily classified according to their dimension. Point defects are zero-dimensional defects, i.e., involve only a single atom which is surrounded by an otherwise perfect lattice. Nonetheless, their presence may affect the properties of the neighboring atoms which, in turn, through elastic interactions affect a sizable region of the lattice around the defect. Although it is possible to induce their formation, point defects are typically intrinsic to the material, that is, they are present since the solid was formed without any specific external origin. They also occur naturally through thermal vibrations. Point defects consist of vacancies, interstitial atoms, and impurities. A vacancy is simply a vacant lattice site which would be otherwise occupied in a perfect crystal. An interstitial is an atom placed in between the normal lattice positions without disturbing the structure of the surrounding lattice. An interstitial may be of the same type of atoms composing the lattice or chemically dissimilar, i.e., an impurity atom. In fact, due to fundamental limitations in the purification methods, materials are never 100% pure. Impurities are often added to materials to improve their properties. Typical examples are carbon impurities added in iron to create steel which is known for being stronger than iron. Similarly, boron impurities added to silicon significantly change its electrical properties.

Along with point defects, there are also defects that are extended in space. One-dimensional defects are called dislocations. Dislocation lines are characterized by a sudden discontinuity in the regular ordering of atoms. A high density of dislocations is typically present in materials, and they play a major role in influencing their mechanical properties. Two-dimensional defects are surfaces, such as the

external surface, grain boundaries, interfaces between different crystallites in a polycrystalline material, and stacking faults, which arise from mismatching in the stacking of crystallographic planes. Three-dimensional defects can change the crystal pattern over a finite volume. Among the most common three-dimensional defects, it is worth mentioning second-phase particles, i.e., non-soluble impurities, which in large amounts can coalesce into precipitates and large voids.

Crystalline materials typically contain large concentrations of more than one of these kinds of defects. It is important to note, however, that in semiconductors even very low densities (< ppm) of defects may have important effects on the electrical and optical properties of the material. The aforementioned different types of defects are known to be able to affect each other's properties and presence. For instance, the presence of stacking faults in a crystal may induce the formation of vacancy defects.

In this chapter, particular emphasis will be on vacancy-type defects in crystalline solids, since the PAS techniques are sensitive to this kind of defects (for conciseness, hereafter we will simply refer to "vacancy-type defects" as "defects" unless otherwise indicated). In order to fully understand the role of defects on the properties of a material, the following steps need to be undertaken:

1. Identify the nature of the defects that are present.
2. Quantitatively measure the concentration of defects.
3. Explore the effects of the defects on the properties of the material.

As described in this chapter, the strength of the existing PAS techniques lies in that they can deliver a range of information which, when combined together, allows, at least in principle, to achieve these targets.

---

## Background of Positron Annihilation Spectroscopy

### Brief History of Positron Annihilation

Dirac (1928) first predicted the existence of the positron in 1928, and just 4 years later, it was experimentally observed by Anderson (1933). When a positron meets an electron, they annihilate each other, and two  $\gamma$ -rays are emitted in opposite directions with an energy of 511 keV each, corresponding to the rest mass energy of each of the two particles. PAS takes advantage precisely of the emission of these annihilation  $\gamma$ -rays to study the electronic and atomic environment of the material where the positron annihilates. Positron-electron annihilation was extensively investigated in the 1940s and 1950s when the first experimental techniques were developed (Bell and Graham 1953; Beringer and Montgomery 1942; Stewart 1957). At the same time, the first theories of positron annihilation in solids were formulated (Ferrell 1956). In the late 1960s and early 1970s, scientists discovered the phenomenon of positron trapping in lattice defects and were able to correlate positron annihilation measurements with the microscopic properties of materials, such as vacancy formation and changing defect concentrations in metals and alloys, as well as molecular

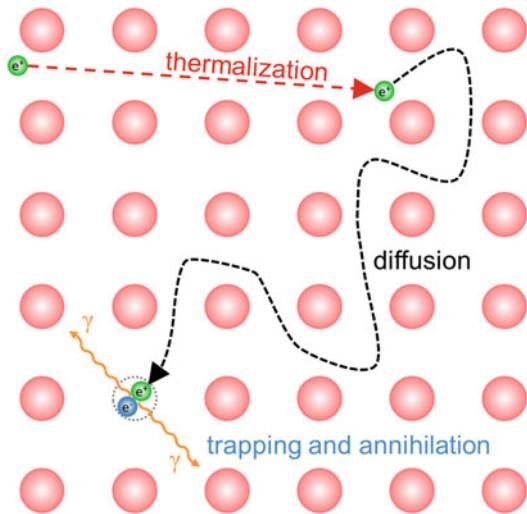
solids (Connors and West 1969; Hodges 1970; Seeger 1974). The successive development of monoenergetic, energy-tunable, slow-positron beams (Lahtinen et al. 1986; Schultz 1988) and the progress in the models of positron annihilation at defects (Manninen and Nieminen 1981; Puska and Nieminen 1994) opened up a new avenue of positron annihilation research on solid surfaces. Hence, it is not surprising that PAS methods have seen rapid growth of applications in materials research ever since the early 1990s (Tuomisto and Makkonen 2013).

## Principles of the Detection of Defects

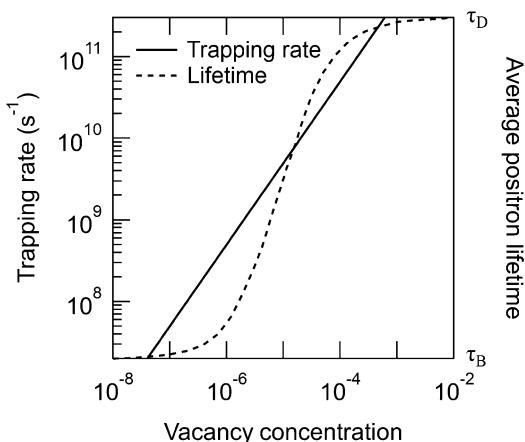
One of the main reasons why the PAS techniques have gained such popularity among all the methods to detect defects is that they can selectively identify vacancy-type defects at the atomic scale with high sensitivity ( $10^{-7}$ ). This is due to the peculiar nature of positrons: they are positively charged and they annihilate with electrons of the crystal lattice. When energetic positrons are emitted from a radioactive source such as  $^{22}\text{Na}$  (see section “[Positron Sources](#)”) and penetrate into a solid material, they quickly lose part of their initial kinetic energy through inelastic collisions with the crystal lattice and the electrons and reach a thermal equilibrium with the environment (Fig. 1). This thermalization process typically lasts just a few ps and after that the thermalized positrons diffuse within the lattice as periodic Bloch states. Due to the strong Coulomb repulsion with the positive ion cores of the crystal lattice, the positron probability density is highest in the interstitial regions. If ions are missing or, in other words, vacancy-type defects are present, the Coulomb repulsion felt by the positrons in the vicinity of these voids is weaker, and the positrons experience these kinds of defects as potential wells. Therefore, positrons may become trapped into bound states in those defects. This transition from a free Bloch state to a localized state at a defect is called positron trapping. These localized positron states are relatively deeply bound: since their binding energy is typically  $>1$  eV (Makkonen and Puska 2007), the thermal escape (detrapping) of positrons from the vacancies can usually be neglected. In general, all open-volume-type defects such as lattice voids, vacancy clusters, dislocations, grain boundaries, and interfaces represent attractive centers for positrons. This is due to the significantly lower local electron density in those kinds of defects compared to the average crystal lattice. As a result, the trapped positrons show a longer lifetime and emit two  $\gamma$ -rays with a somewhat narrower energy width compared to those annihilating in the defect-free lattice. Based upon these principles, PAS can be considered a probe of vacancy-type defects in materials.

The strength of the positron annihilation-based technique lies in the high sensitivity of positrons. The detection range of positrons varies from an atomic ratio of  $10^{-7}$  to  $10^{-3}$ . The sensitivity of positrons is defect dependent, and in particular it is determined by the nature of defects (i.e., voids, dislocations, clusters, etc.) and their concentration in the material (Fig. 2). Figure 2 shows an example of the positron trapping rate (see section “[Trapping Model](#)”) and average lifetime in silicon calculated as a function of vacancy concentration. It is clear that there is a correlation

**Fig. 1** Schematic representation of the different processes that a positron undergoes after implantation in a crystal lattice: thermalization (slowing down), diffusion, trapping at a vacancy defect, and annihilation with an electron



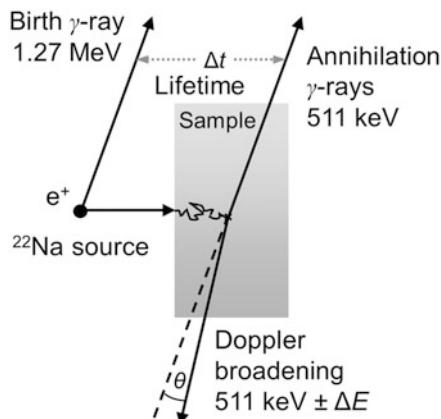
**Fig. 2** Positron trapping rate and average positron lifetime in Si calculated as a function of the vacancy concentration in the material.  $\tau_D$  is the positron lifetime at defects and  $\tau_B$  is the positron lifetime in the bulk. Numerical data are from Krause-Rehberg and Leipner (1997)



between these parameters: the higher the defect concentration, the higher are the positron lifetime and the trapping rate. The high sensitivity to a large variety of defects, defect sizes, and concentrations and the flexibility to examine the bulk and the surface layers of a large variety of materials are the fundamental features that make PAS a unique tool for the micro-characterization of the structure of materials and contributed to its success. In addition, PAS has sound theoretical foundations, since the annihilation properties can be simply calculated from the first principles.

Two of the basic PAS techniques, namely, PALS and DBS, have been efficiently applied to defect studies on a large variety of materials. Both are based on the detection of the two  $\gamma$ -rays that are emitted from the positron-electron annihilation (Fig. 3). Owing to the energy and momentum conservation in the annihilation process, the two  $\gamma$ -rays are emitted in opposite directions, with an angular deviation

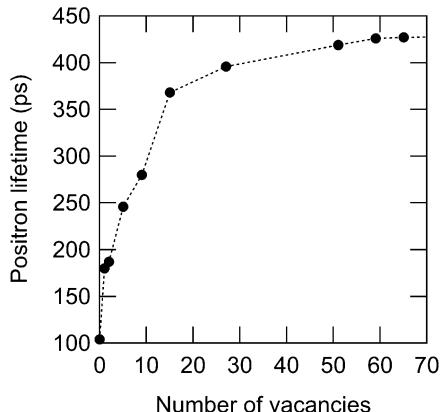
**Fig. 3** Schematic representation of the positron annihilation principles which form the basis for positron annihilation lifetime spectroscopy and Doppler broadening spectroscopy



$\theta$ , and with an energy of 511 keV, plus/minus an energy shift  $\Delta E$ . The angular deviation  $\theta$  from the straight angle between the two photons and the Doppler shift  $\Delta E$  in their energy are due to the finite momentum of the annihilated electrons. Positron annihilation at vacancy-type defects induces changes in the electron momentum distribution. In particular, the momentum distribution originating from valence electron annihilation becomes a little narrower due to the lower electron density. In addition, given the reduced annihilation probability of localized positrons at a vacancy with core electrons, a significant reduction of the high-momentum component in the momentum distribution of the annihilated electrons is expected. The annihilation photons contain all this information on the state of the annihilated positrons and electrons. If those photons can be experimentally detected, the electronic properties and crystal structure of solids can be probed.

In a PALS experiment, the distribution of the time difference between the injection of the positrons into the material and their annihilation is measured. The positron annihilation rate per unit time is proportional to the electron density distribution encountered by the positrons in the material. Thus, the positron lifetime, i.e., the inverse of the positron annihilation rate, which is determined experimentally from a PALS spectrum, yields information about the electron density in the region of the material where the positrons annihilated. In particular, the positron lifetime is inversely proportional to the local electron density. Thus the positrons that become trapped and annihilate in defects (i.e., in a region with lower electron density than average) have longer lifetimes than those annihilating in the defect-free lattice. In other words, PALS measurements can probe the presence of defects in the materials. Positrons can measure an extended range of lifetimes and, therefore, are sensitive to a wide variety of defect sizes. Positron lifetimes in the bulk of most metals range from about 100 to 150 ps, while lifetimes of positron trapped in monovacancies and dislocations are some ~50% longer. Positron lifetimes in vacancy clusters typically range between this value and ~450 ps and increase with the size of the cluster (Fig. 4).

**Fig. 4** Positron lifetime of vacancy clusters in Fe as a function of cluster size obtained from first-principle calculations. The line is drawn just to guide the eye. Numerical data are from Ohkubo et al. (2003)



DBS measurements yield the probability distribution of the deviations in  $\Delta E$ . Those measurements provide detailed information about the electron momentum states and, therefore, the chemical nature of defects from the region in which the positrons annihilated. As described in detail in section “[Data Analysis](#),” two parameters called  $S$  and  $W$  are usually defined. They are sensitive to the fraction of low-momentum annihilation events (typically with valence electrons) near the center of the spectrum and high-momentum annihilation events (usually with core electrons) in the wings of the spectrum, respectively.

## Trapping Model

As mentioned above, positron trapping occurs when a delocalized positron state becomes bound to a defect as a localized state. For a positron to get trapped at a vacancy, the trapping process must be fast enough to compete with annihilation. During a typical PAS measurement, where positrons emitted by a radioactive source (see section “[Positron Sources](#)”) are directly injected into a sample, only one positron at a time is present in the specimen. The positron trapping and annihilation that occurs in this situation can be described by the so-called trapping model (West 1979), i.e., a set of relatively simple kinetic rate equations. Because the slowing down and thermalization processes are much faster than trapping and annihilation, the model is based on the following assumptions:

1. At time 0, all positrons are in a free state.
2. The positron trapping rate  $\kappa_D$  into a defect  $D$  is proportional to the defect concentration  $c_D$ :  $\kappa_D = \mu_D c_D / N_{\text{at}}$ , where  $N_{\text{at}}$  is the atomic density of the host lattice. This defines the trapping coefficient  $\mu_D$ , i.e., the specific trapping rate for a unit concentration of defects, which depends on the defect and the host lattice.
3. A trapped positron may escape from the trap. The detrapping rate is denoted by  $\delta_D$ .

Let's assume that there are  $N$  different defects in the sample. The probability of the positron to be in a free state is denoted by  $n_B$  and that of being in a trapped state at defect  $D$  by  $n_D$ . Further,  $\lambda_B$  and  $\lambda_D$  are the annihilation rates for free and trapped positrons, respectively. Given that the positron lifetime is defined as the inverse of the positron annihilation rate,  $\tau_B = 1/\lambda_B$  and  $\tau_D = 1/\lambda_D$  are the lifetimes for free and trapped positrons, respectively. A set of linear differential equations can be written in the form:

$$\frac{dn_B}{dt} = -\left(\lambda_B + \sum_D \kappa_D\right)n_B + \sum_D \delta_D n_D, \quad (1)$$

$$\frac{dn_D}{dt} = \kappa_D n_B - (\lambda_D + \delta_D) n_D, \quad (2)$$

where the summations are over all defects  $D$  ( $D = 1, \dots, N$ ). Typically,  $\delta_D$  is negligible except for weakly bound states at high temperature.

In the simplest case where there is only one type of defect in the sample, these kinetic equations can be rewritten as:

$$\frac{dn_B}{dt} = -(\lambda_B + \kappa_D)n_B + f, \quad (3)$$

$$\frac{dn_D}{dt} = \kappa_D n_B - \lambda_D n_D, \quad (4)$$

where  $f$  is a source term, i.e., the number of detrapped positrons reentering the system. By applying the initial conditions  $n_B(t=0) = 1$  and  $n_D(t=0) = 0$ , this set of equations can be solved, and the probability a positron to be in the sample at time  $t$  is given by:

$$n(t) = n_B(t) + n_D(t) = \sum_i I_i e^{\lambda_i t}. \quad (5)$$

By differentiating Eq. (5) with respect to time, the probability of positron annihilation in the time interval from  $t$  to  $t+dt$  can be obtained.  $-dn(t)/dt$  is precisely what is measured in a positron lifetime spectrum. In this case, the positrons in a free state are assumed to annihilate in the bulk  $B$  of the material, whereas the trapped positrons annihilate in the defects  $D$ . Hence, the experimental spectrum is composed of the sum of two components, each represented by an exponential decay function:

$$\frac{dn}{dt} = -n_0 \left( \frac{I_1}{\tau_1} e^{-t/\tau_1} + \frac{I_2}{\tau_2} e^{-t/\tau_2} \right). \quad (6)$$

The relationships between the trapping rate  $\kappa_D$ , the positron lifetime in the bulk  $\tau_B$ , the positron lifetime at the defects  $\tau_D$ , and the two-component fitting parameters  $(\tau_1, \tau_2, I_1, I_2)$  in Eq. (6) are given by:

$$1/\tau_1 = 1/\tau_B + \kappa_D, \quad (7)$$

$$\tau_2 = \tau_D, \quad (8)$$

$$I_2 = 1 - I_1 = \kappa_D / (\kappa_D + \lambda_B + \lambda_D). \quad (9)$$

These expressions are very useful because they directly relate experimentally determined time-averaged quantities with intrinsic properties of the material such as the positron lifetime and the trapping coefficient. Parameters describing the Doppler broadening of the annihilation radiation can also be associated with these properties as those parameters measure the superposition of the annihilation over all positron states.

An average positron lifetime can also be defined:

$$\tau_{av} = \eta_B \tau_B + \eta_D \tau_D = I_1 \tau_1 + I_2 \tau_2, \quad (10)$$

where  $\eta_B$  and  $\eta_D$  are the annihilation fractions in the bulk and at the defects, respectively, and  $\eta_B + \eta_D = I_1 + I_2 = 1$ . The importance of the average positron lifetime is manifested by Eq. (10) which directly relates the measured PALS spectrum to the trapping model. It can be estimated with high statistical accuracy owing to the large number of annihilation events and the stability of the PALS systems. Therefore, it is often used as a key parameter in PALS analysis. The annihilation fractions in the bulk and at the defects are, respectively, given in this case by:

$$\eta_B = \lambda_B / (\lambda_B + \kappa_D) \quad (11)$$

$$\eta_D = \kappa_D / (\lambda_B + \kappa_D) \quad (12)$$

The trapping rate is given in terms of the experimentally determined parameters by:

$$\kappa_D = \lambda_B \frac{\tau_{av} - \tau_B}{\tau_D - \tau_{av}}. \quad (13)$$

The defect concentration can be determined from Eq. (13) provided that the bulk lifetime  $\tau_B$  is known. This can be achieved, for instance, by measuring beforehand the corresponding defect-free material, or a reference sample, where no positron trapping at defects occurs.

The experimental estimation of the values of the trapping coefficient through the trapping model is important because they are needed to extract defect concentrations and formation energies from the positron lifetime measurements. Nonetheless, the determination of absolute defect concentrations based on positron experiments is somewhat difficult because it directly depends on the knowledge of the trapping coefficient. However, even when the trapping coefficient is unknown, the comparison of defect concentrations, for example, between two different samples, of course provides accurate proportions.

The trapping coefficient for vacancies in metals is typically of the order of  $\mu_D = 10^{14}\text{--}10^{15} \text{ s}^{-1}$  (West 1979). The magnitude of the trapping coefficient is about five orders of magnitude larger than the typical positron annihilation rates ( $\lambda_B \approx 5 \times 10^9 \text{ s}^{-1}$ ). It must be noted that the positron technique only detects vacancies and other defects in their negatively charged and neutral states. The trapping coefficient at neutral vacancies is typically  $\mu_D = 10^{14}\text{--}10^{15} \text{ s}^{-1}$  and is temperature-independent. The trapping coefficient at negatively charged vacancies normally amounts to  $\mu_D = 10^{15}\text{--}10^{16} \text{ s}^{-1}$  at room temperature and increases as the temperature decreases with a  $T^{-1/2}$  dependence. Owing to the Coulomb repulsion, positrons are prevented from being trapped at positively charged vacancies within the short positron lifetime of a few hundred ps (Puska et al. 1990). For more details about the trapping model, please see, e.g., Bertolaccini et al. (1971) and Brandt (1974).

## Positronium Formation in Molecular Systems

It is important to mention here that a special feature of positron interactions with materials with low enough electron density is that a positron can bind to an electron to form a metastable bound state called positronium (Ps) (Charlton and Humberston 2001). Ps is formed in materials without free electrons, i.e., insulating materials such as polymers and inorganic glasses, as well as liquids and gases. Ps cannot form in the bulk of a metal or semiconductor because the electron density is too large. Ps is an overall neutral, hydrogen-like exotic atom with a binding energy of 6.8 eV in vacuum and a diameter of 1.06 Å in its ground state in vacuum. Ps can form in either the singlet (antiparallel spins, para-positronium, p-Ps) or triplet (parallel spins, ortho-positronium, o-Ps) state depending on the relative spins of the two composing particles. Ps is an unstable system with a mean lifetime of 125 ps (p-Ps) or 142 ns (o-Ps) in vacuum, and the two particles annihilate each other to produce either two or three  $\gamma$ -rays, respectively. Given its very short lifetime, p-Ps annihilates without interacting with the host material. On the other hand, given that o-Ps forms in the material and there are electrons in the surroundings, the positron in the pair likely annihilates with another electron (with opposite spin direction) nearby other than its partner and emits two photons instead of three. This phenomenon, which is called pick-off annihilation of o-Ps, is prevalent in condensed matter (Jean et al. 2003).

The properties of Ps become very useful especially when the PAS techniques are applied to study free volume (i.e., empty microscopic space that exists between molecules) in molecular systems and amorphous solids, e.g., polymers, or voids and pores in porous materials such as zeolites (Mogensen 1995). It is important to note that the principles, methods, and data analysis techniques become somewhat different when Ps is involved. Although it is out of the scope of this chapter to provide a comprehensive description of the Ps-based techniques, it is worth mentioning here that the o-Ps lifetime in condensed matter is reduced by the interaction with molecular electrons during collisions with the pore surface with respect to its value in vacuum. Hence, the reduced o-Ps lifetime is correlated to the average pore size or

free volume, while the intensity of this component relates to the amount of such defects (Schrader and Jean 1983). This principle forms the physical basis for probing the structure of this kind of systems through *positronium* annihilation lifetime spectroscopy.

The relationship between the experimentally determined o-Ps lifetime  $\tau_3$  (ns) and the average hole radius  $r$  (nm) is determined through quantum mechanical calculations (Tao 1972) which model free volume as a spherical potential well surrounded by an electron layer of thickness  $\Delta r$ :

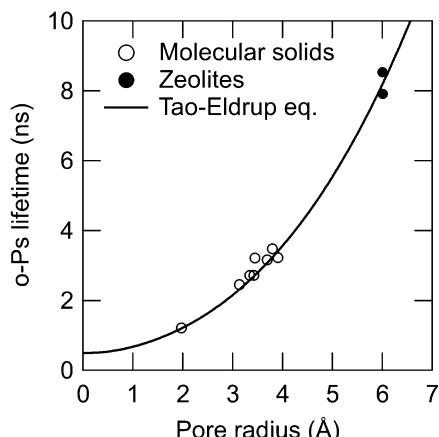
$$\frac{1}{\tau_3} = 2 \left[ 1 - \frac{r}{r_0} + \frac{1}{2\pi} \sin \left( \frac{2\pi r}{r_0} \right) \right]. \quad (14)$$

In Eq. (14),  $r_0 = r + \Delta r$ , where  $\Delta r = 0.166$  nm is assumed to be the same for all materials and has been experimentally determined by fitting  $\tau_3$  values to known hole sizes, such as those of zeolites. Equation (14) represents the so-called Tao-Eldrup equation (Tao 1972; Eldrup et al. 1981). The free-volume size  $V_f$  is, in turn, given by:

$$V_f = \frac{4}{3}\pi r^3. \quad (15)$$

It is important to note that there is a minimum threshold for the hole size that can be detected by Ps annihilation lifetime spectroscopy: this corresponds to 1.3 Å. Although in the Tao-Eldrup model holes of irregular shape are treated as spheres for simplicity, this model can be applied to other pore shapes (cubic, cuboidal, cylindrical, etc.) as well. Figure 5 shows the relationship between pick-off o-Ps lifetime and pore radius according to the Tao-Eldrup equation in comparison to some experimental data for vacancies in molecular solids and zeolites. The model fits very well the measured data.

**Fig. 5** Measured pick-off o-Ps lifetime of vacancies in selected molecular solids and zeolites as a function of the pore radius. The curve shows the fit using the Tao-Eldrup equation. Numerical data for molecular solids are from Eldrup et al. (1981) and for zeolites from Nakanishi et al. (1988)

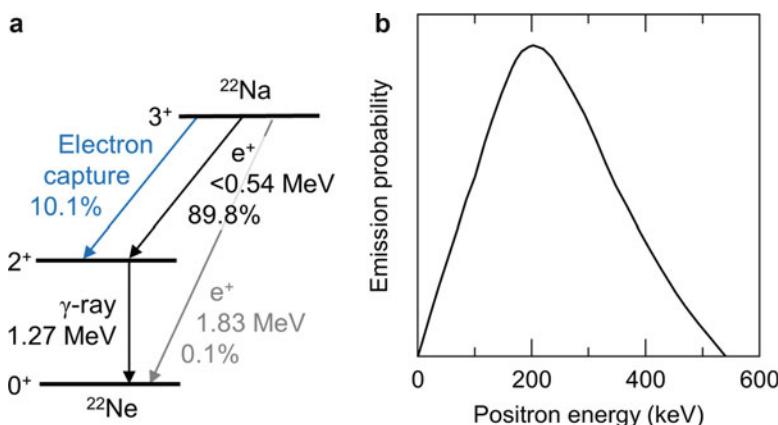


## Experimental Techniques

In this section, the most extensively and commonly used methods in defect studies using positrons, namely, PALS, DBS, and VEPAS, are described and discussed. An overview of the experimental details, the data analysis procedure, and the information that can be obtained from those techniques is presented. More comprehensive reviews of the PAS methodology and other positron techniques can be found elsewhere (e.g., Charlton and Humberston 2001; Coleman 2000; Gidley et al. 2006; Jean et al. 2003, 2013; Krause-Rehberg and Leipner 1999; Mallon and Schrader 2003; Mogensen 1995; Schultz and Lynn 1988; Siegel 1980; Singh 2016; Tuomisto and Makkonen 2013).

### Positron Sources

Although antimatter seems to be quite rare in the visible universe, positrons can be generated in many ways. Positrons are quite readily available, for instance, from radioactive isotopes such as  $^{40}\text{K}$ ,  $^{22}\text{Na}$ ,  $^{18}\text{F}$ ,  $\text{Cu}^{64}$ , and  $\text{Co}^{58}$  which are commercially produced. The most common radioisotope used in positron experiments is  $^{22}\text{Na}$  given its half-life of 2.6 years which allows the use of the same source for a reasonable time frame (5–10 years).  $^{22}\text{Na}$  decays to an excited  $^{22}\text{Ne}$  nucleus via  $\beta^+$  decay with the emission of a neutrino and a positron (Fig. 6a). The excited  $^{22}\text{Ne}$  nucleus, in turn, rapidly decays with the emission of a 1.27 MeV  $\gamma$ -ray. In this process the positrons are emitted with a continuous and wide energy spectrum up to



**Fig. 6** (a) Schematic diagram of the decay processes of  $^{22}\text{Na}$ . The main channel is  $\beta^+$  decay (89.8%) from  $^{22}\text{Na}$  to excited  $^{22}\text{Ne}$  with an energy spectrum of up to 540 keV. A  $\gamma$ -ray with an energy of 1.27 MeV is subsequently (about 3 ps later) emitted upon decay of the excited  $^{22}\text{Ne}$  to the ground state. Electron capture takes place with a probability of 10.1% and there is no positron emission. Direct  $\beta^+$  decay of  $^{22}\text{Na}$  to the ground state of  $^{22}\text{Ne}$  may also occur with a very low probability ( $<0.1\%$ ). (b) Approximate shape of the energy distribution of positrons emitted from  $^{22}\text{Na}$

540 keV (Fig. 6b). Such high energies ensure that the positrons are implanted deep into the sample, typically up to 300  $\mu\text{m}$  in metals.

As  $^{22}\text{Na}$  typically comes as NaCl in the form of a water solution, the positron source usually needs to be prepared ahead of the experiment. A drop of NaCl solution is deposited in between two thin foils (made of polymers such as Kapton and Mylar or metals such as Al and Ni) which are sealed to each other after the water has evaporated. The typical activity of a source prepared in such a way is 0.3–1.0 MBq. Given that the positron source is enclosed between these two foils, it can be reused for several different experiments. The source needs to be as thin as possible in order to minimize the number of positrons annihilating in the films enclosing the source itself. For this purpose, foils with a thickness of about 1–2  $\mu\text{m}$  are typically used. Sometimes some NaCl is deposited directly onto the surface of one of the two specimens, and then another identical sample is laid on top of it. Although this may be the solution of choice when the measurements are carried out over a wide range of temperatures, in this case the source cannot be reused, and contamination of the sample represents a significant issue.

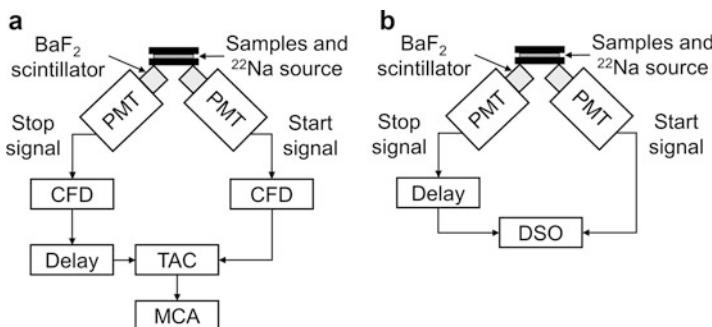
The energy spectrum of positrons emitted from  $^{22}\text{Na}$  may be very wide (Fig. 6b), but the average kinetic energy is around 200 keV. The stopping profile of energetic positrons emitted by a radioactive source such as  $^{22}\text{Na}$  follows a Makhovian distribution (see section “[Variable Energy Positron Annihilation Spectroscopy](#)”), i.e., it is exponential as a function of depth (Brandt and Paulin 1977). For this isotope in particular, the characteristic penetration depth ranges from 0.1 to 0.3 mm. This means that positrons emitted from the source and directly implanted into the material can penetrate deep into the sample and probe the bulk properties of a solid. As such, this approach is unsuitable to study surface layers such as thin films or coated substrates. This issue can be overcome by developing variable-energy and mono-energetic positron beams starting from radioactive sources (see section “[Variable Energy Positron Annihilation Spectroscopy](#)” for more details).

## Positron Annihilation Lifetime Spectroscopy

### Experimental Details

The typical geometry of a PALS spectrometer is shown in Fig. 7. A PALS experiment is carried out by placing the positron source in between two samples, typically 1–2-mm-thick identical replicas of the material of interest. The positron lifetime in the material can be estimated by measuring the time difference between the birth  $\gamma$ -quantum with an energy of 1.27 MeV (start signal) and one of the two 511-keV annihilation photons (stop signal). This is achieved by detecting those photons using two  $\gamma$ -ray scintillation detectors (such as BaF<sub>2</sub>) coupled to photomultiplier tubes (PMTs). The energy spectrum of each detector is tuned in order to ensure that only one of the two  $\gamma$ -quanta (either the start or the stop signal) is detected.

In laboratories around the world, both analog (Fig. 7a) and digital (Fig. 7b) PALS experimental setups have been widely employed, although the latter have overwhelmingly become more prevalent in recent times. In an analog system, the signals



**Fig. 7** Schematic representation of (a) an analog and (b) a digital PALS spectrometer. PMT stands for photomultiplier tube, CFD for constant fraction discriminator, “delay” indicates a time delay module, TAC is a time-to-amplitude converter, MCA means multichannel analyzer, and DSO represents a digital storage oscilloscope

from the scintillation detectors coupled to the photomultipliers are delivered to a set of nuclear instrumentation for processing and optimization. Electronics modules typically include:

- A constant fraction discriminator (CFD) enables one to select from the energy spectrum only those photons detected in the desired energy ranges (i.e., those around the 511-keV and the 1.27-MeV photo-peaks).
- A time-to-amplitude converter (TAC) converts the time intervals between the detected photons to pulses with an amplitude proportional to that time.
- A multichannel analyzer (MCA) collects and digitizes the pulses using an internal analog-to-digital converter and stores the signal information into different channels based on their amplitude or time (typical resolutions are 10–20 ps/channel).

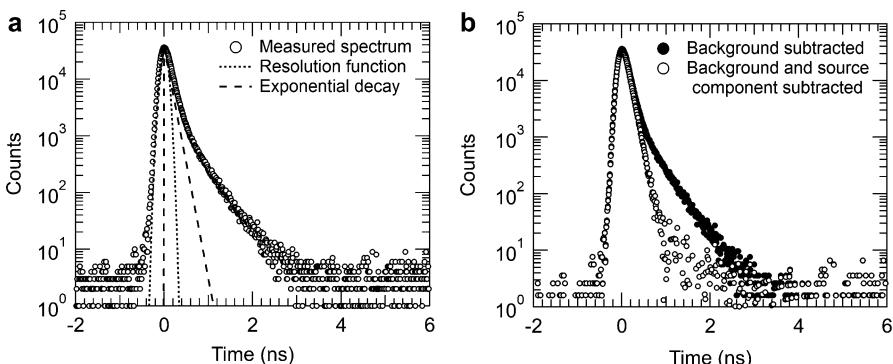
The MCA is usually interfaced with a PC for data visualization and analysis using existing software. The data collected this way result in a histogram of the number of annihilation events as a function of the time difference between the start signal (i.e., positron emission or implantation into the sample) and the stop signal (i.e., positron annihilation). This histogram is called a positron annihilation lifetime spectrum as, in other words, it contains information about the lifetimes of the positrons in the sample. Recently, digital PALS systems have become increasingly more common because they allow one to obtain a lifetime spectrum simply through direct digitization, measurement, storage, and display of the signals, for example, with a digital storage oscilloscope (DSO) (see, e.g., Saito et al. 2002) (Fig. 7b). The advantage of DSOs is that they support post-measurement signal analysis using pre-installed software. This represents a significant simplification of the entire PALS system and reduces the overall measurement and analysis time. Another important improvement of digital systems is that they possess a superior instrumental time resolution (see section “[Data Analysis and Information from PALS Spectra](#)”).

## Data Analysis and Information from PALS Spectra

A PALS spectrum is a lifetime histogram of all the annihilation events of the implanted positrons. In order to achieve a sufficiently high statistical accuracy, a collected lifetime spectrum typically contains  $\sim 1 \times 10^6$  annihilation events. Figure 8a shows an example of a “raw” lifetime spectrum for defect-free, high-purity Fe obtained with a typical PALS spectrometer. The presence of exponential decay functions, as mentioned in section “[Trapping Model](#),” is manifested in Fig. 8a. However, there are clearly other important features in the spectrum that are important to note before analyzing it.

First, there is always inevitably some level of background noise in the spectrum. In principle, given that this is a coincidence measurement, there should be no random background noise. Nonetheless, positrons from a radioisotope are not emitted in a deterministic way; in other words, the decay of a given nucleus is an entirely random event. Hence, despite the fact that the source activity is very low, say, for instance,  $\sim 1$  MBq, which corresponds to an average time difference of  $\sim 1\ \mu\text{s}$  between two consecutive decays, two positrons might actually be emitted almost simultaneously. This results in false coincidence counts which appear as a flat background in the spectrum. In PALS measurements, the background level is determined only by the source activity ([Knoll 2000](#)). For this reason, the signal-to-noise ratio can be improved by reducing the source activity and correspondingly increasing the measurement time. Figure 8b shows the same spectrum of panel (a) after the background has been subtracted.

Second, the detectors used for the PALS experiments possess an inherently finite time resolution. The extent of the time resolution depends on the nature and size of the scintillator, as well as the optimization of the photomultiplier tube settings (e.g., the supply voltage) ([Ackermann et al. 2015](#)). As a consequence, the measured spectra are inevitably the result of the convolution of the real lifetime spectra over the overall distribution of the instrumental time resolution. As is apparent from



**Fig. 8** (a) A “raw” positron lifetime spectrum of defect-free, high-purity Fe obtained with a typical PALS spectrometer. The instrumental resolution of the system and the exponential decay component with a lifetime of  $\sim 107\ \text{ps}$  arising from positrons annihilating in the bulk are also shown. (b) The same spectrum of (a) after background and subsequent source component subtraction

Fig. 8a, this is a non-negligible effect because the shape of the peak centered at around  $t = 0$  is not sharp as one would ideally expect but is rather smoothed out by the width of the time resolution function. For this reason, the measured lifetime spectrum is not simply given by the sum of exponential functions (see Eq. 5) but needs to be fitted with the sum of exponential functions convoluted over the instrumental time resolution function. The detailed shape of the time resolution function is a priori unknown. However, it can be estimated, for instance, by measuring the lifetime spectrum of a defect-free material whose positron lifetime is considered to be well known (such as metallic or glass reference samples). Deconvolution of this kind of spectrum during the data analysis process using existing software (see below) allows one to obtain the parameters of the resolution function. Typically, the time resolution function can be assumed to be Gaussian in shape, and one or two Gaussian distributions are used to fit it. If the time resolution function is indeed Gaussian in shape, it means that the instruments used for the PALS system are merely a source of statistical uncertainty and do not introduce any systematic error in the measurements. In addition, data analysis remains relatively simple as the convolution with Gaussian functions does not alter the shape of the exponential decay components. Figure 8a shows the resolution function of a typical PALS spectrometer obtained from the fitting of the lifetime spectrum present in the same figure. In this case the resolution function is a Gaussian distribution with a full width at half maximum (FWHM) of  $\sim 175$  ps. However, resolutions achieved in experimental setups can be as low as 140 ps (Bečvář et al. 2008). A resolution function with a FWHM of this magnitude allows successful determination of the exponential decay components of the lifetime spectrum as long as their lifetime is larger than 30 ps or so, their relative intensity is high enough, and the time resolution of the spectrum is at least 10 ps/channel. For this same reason, on the other hand, very short lifetimes with low intensity cannot be reliably extracted from the spectrum. In the spectrum of Fig. 8a, in particular, a component with a lifetime of  $\sim 107$  ps (corresponding to positrons annihilating in the Fe bulk) is present.

Third, the positron source is typically enclosed between two identical polymer or metallic foils. Therefore, a fraction of the positrons inevitably annihilates in those foils before implantation in the samples. This results in an additional component in the measured lifetime spectrum that needs to be removed before proceeding with the data analysis. Figure 8b shows the same spectrum of panel (a) after the background and source component have been subtracted. Usually, the source component can also be represented by one exponential decay function, which makes it quite easy to fit. The lifetime of the source component depends on the nature of the foils enclosing the source, while its relative intensity is determined by the average atomic number of the sample and, most importantly, the thickness of the foils (Bertolaccini and Zappa 1967). For example, in the case of 0.7- $\mu\text{m}$ -thick Kapton-sealed  $^{22}\text{Na}$ , the source component typically has only a single lifetime of  $\sim 380$  ps and a relative intensity of around 12–17% in common metals. The presence of a single lifetime in Kapton is an advantage compared to other materials (which may present more than one lifetime), as it simplifies the resolution of other lifetime components in the spectrum during the analysis process. To determine the lifetime and relative intensity of the source

component, typically one needs to measure high-purity, defect-free reference samples with high surface quality to ensure that no positron trapping at defects occurs. The extent of the relative intensity of the source component is not an issue, as it is subtracted before the data is analyzed further. Moreover, given that PALS is an inherently comparative technique, i.e., where changes in the properties of different samples are compared, the analysis of lifetime spectra is not affected by the presence of a source component as long as all samples have been measured with the same setup and the same source.

A PALS spectrum is a combination of several exponential components. Therefore, mathematical modeling can be used to fit a number of exponential decay functions through convolution with the resolution function. Given the intrinsic characteristics of the spectrum outlined above, only a limited number of components can be extracted from the spectrum. Typically, up to three lifetimes can be distinguished in a spectrum (besides the source component), provided that they differ in magnitude by at least some 30% or so. The spectrum is usually fitted using an adapted version of Eq. (6) that includes the desired number of exponential decay components after background subtraction and source correction. The fitting parameters include the intensities  $I_i$  and the annihilation rates  $\lambda_i$  of the various lifetime components, as well as the FWHM of the resolution function. The fitting process is carried out by least squares methods, and the variance of the fit describes how well the model fits the observations. Typical data analysis programs that follow this methodology are PositronFit (Kirkegaard and Eldrup 1972), PALSfit (Olsen et al. 2007), RESOLUTION (Kirkegaard and Eldrup 1974), and LT (Kansy 1996).

Analysis of a defect-containing sample commonly reveals the presence of at least two lifetime components ( $\tau_1, \tau_2$ ) in the spectrum (see Eq. 6), because the positrons annihilate in no less than two different states. The first component resembles the delocalized states in the lattice, while the second component (and other components, if present) is expected to originate from localized states at defects. The shortest component ( $\tau_1$ ) corresponds to the so-called reduced bulk lifetime, i.e., the apparent positron lifetime in the lattice which is somewhat shortened by the trapping process in comparison to the bulk lifetime of the defect-free material (see Eq. 7). The longest component ( $\tau_2$ ), instead, directly reflects the lifetime of the specific kind of defects present in the material (see Eq. 8). The trapping rate of the defects in question and the defect concentration can be determined from the experimentally obtained parameters ( $\tau_1, \tau_2, I_1, I_2$ ) using the trapping model (see section “[Trapping Model](#)”).

## Doppler Broadening Spectroscopy

### Experimental Details

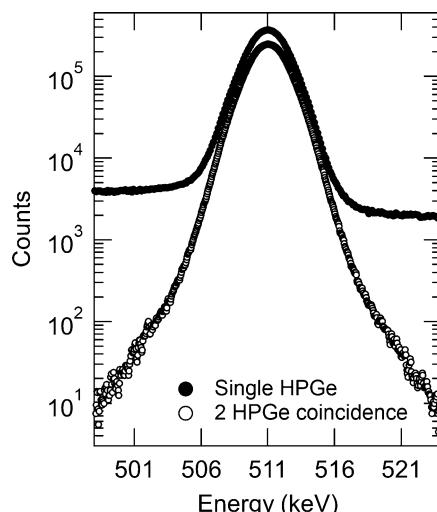
An experimental setup for DBS measurements employs high-energy-resolution  $\gamma$ -detectors, typically high-purity Ge (HPGe) detectors, to detect the annihilation photons. In an analog DBS spectrometer, the detector signals are supplied to a MCA, which collects the information into different channels based on the energy. The MCA, in turn, is interfaced to a PC. Digital DBS systems have recently also

become available and replace the aforementioned instrumentation through direct digitization, measurement, storage, and display of the signals. The data collected in this manner result in an energy spectrum of the  $2\gamma$ -annihilation events in a window centered around the 511 keV peak.

The annihilation events that give rise to the  $2\gamma$ -radiation stem principally from free positron annihilation. In the rest frame, this annihilation radiation consists of  $2\gamma$ -rays of exactly 511 keV each. However, in the laboratory frame, the measured energy of the annihilation radiation may be Doppler shifted owing to the center-of-mass motion of the electron-positron pair. This energy shift is given by  $\Delta E = cp_L/2$ , where  $p_L$  is the longitudinal momentum component of the pair in the direction of the 511 keV annihilation photon emission. As a result, the measured energy spectrum of the  $2\gamma$ -annihilation radiation is Doppler broadened and in some cases Doppler-shifted. The broadening of the 511 keV peak is mostly determined by the electron momentum distribution. Hence, the shape of the 511 keV annihilation peak contains information about the one-dimensional momentum distribution of the annihilating electron-positron pairs. The DBS technique takes advantage of the fact that a pair momentum with a kinetic energy of just a few eV can produce energy shifts of the order of keV in the gamma rays emitted along the pair momentum.

Figure 9 shows the 511 keV peak of a typical Doppler broadening spectrum for defect-free, high-purity Fe measured with a single HPGe detector. The typical resolution of a HPGe detector is around 1–1.5 keV at 511 keV, which is comparable in size to the 2–3 keV width of the peak. The shape of the measured spectrum is, therefore, inevitably convoluted over the resolution function of the detector. Hence, the shape of the measured spectrum strongly depends on the detector resolution. It is manifested in Fig. 9 that the peak-to-background ratio is quite low, since it amounts to just about  $10^2$ . For this reasons sections of the annihilation peak are often

**Fig. 9** Typical Doppler broadening spectrum of defect-free, high-purity Fe measured with a single HPGe detector and typical coincidence Doppler broadening spectrum of the same sample obtained with two HPGe detectors in coincidence



integrated, and various shape parameters are defined to characterize the spectrum properties (see section “[Data Analysis](#)”).

In order to significantly improve the peak-to-background ratio and the energy resolution of the DBS spectra, the annihilation radiation can be measured simultaneously using two detectors and by detecting the  $\gamma$ -ray pairs in coincidence (Asoka-Kumar et al. 1996). This improved version of the DBS technique is often referred to as coincidence Doppler broadening (CDB) spectroscopy. In order to achieve the coincidence in a CDB experiment, a number of conditions need to be met:

1. Collinearity between the two detectors and the sample: the two detectors need to be arranged in a collinear geometry with the sample because the annihilation  $\gamma$ -photon pairs are emitted in almost exactly opposite directions.
2. Time coincidence: only those  $\gamma$ -ray pairs that are emitted in coincidence are selected in order to ensure that the photons originate from the same annihilation event.
3. HPGe detectors: both detectors are high-energy-resolution HPGe detectors.

If only the first two conditions are satisfied, e.g., when just one HPGe detector is used in coincidence with a scintillator detector of poorer energy resolution (this is sometimes the choice as the latter is significantly cheaper than a Ge detector), the peak-to-background ratio of the CDB spectrum increases to about  $10^4$ . If the third condition is also met, that ratio improves even further to around  $10^5$ . In fact, the third condition requires that the sum of the energies  $E_1$  and  $E_2$  of the detected coincidence  $\gamma$ -ray pairs satisfies Eq. (16):

$$E_1 + E_2 = 2m_0c^2 - E_{+,-} \approx 1.022 \text{ MeV}. \quad (16)$$

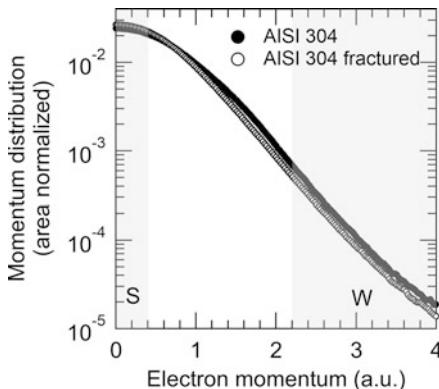
Equation (16) ensures that the energy of the two photons corresponds to the rest mass of the annihilating positron and electron (511 keV each) reduced by the binding energy of the pair ( $E_{+,-}$ ).  $E_{+,-}$  is of the order of  $\sim$ eV that is negligible compared to the rest mass of the positron and electron and can be neglected in Eq. (16). Figure 9 shows the region around the annihilation peak of a typical CDB spectrum for the same defect-free, high-purity Fe sample measured with two HPGe detectors in coincidence. The drastic improvement in the peak-to-background ratio is evident.

CDB measurements generate a two-dimensional matrix where the number of coincident annihilation events is given as a function of the energy measured by the first and the second detector. By integrating the data on one diagonal of the 2D matrix, the CDB spectrum can be plotted, whereas the other diagonal represents the overall resolution function of the spectrometer. The latter is, in fact, different from the resolution of each individual detector, precisely by a factor of  $1/\sqrt{2}$ . For more information and details about the CDB technique, the interested reader may refer to Asoka-Kumar et al. (1996).

## Data Analysis

The figure of interest in CDB spectra is the shape of the 511 keV annihilation peak. CDB spectra are usually normalized to the total area under the peak, folded along the

**Fig. 10** Folded experimental coincidence Doppler broadening spectrum of AISI 304 stainless steel before and after plastic deformation and fracturing. Typical integration windows for the *S* and *W* parameters are shown by the shaded areas



511 keV line, and adjacent counts are summed together. This improves the statistics of the spectrum, especially in the wing regions where the number of annihilation events is reduced. In doing so, the 511 keV line becomes the 0-momentum line in the folded spectrum. Typical area normalized and folded spectra for AISI 304 stainless steel before and after plastic deformation and fracturing are shown in Fig. 10. In order to extract useful information from a CDB spectrum and to compare various spectra, two parameters are defined to quantify the annihilation fraction in different sections of the spectrum. They are described based on the designation of a narrow energy window centered around 511 keV. The so-called *S* parameter is the ratio of the annihilation events in this central window to the total number of counts in the peak (Fig. 10). This parameter reflects the relative fraction of low-momentum annihilation events. Similarly, the *W* parameter is defined as the ratio of the integrated annihilation events in a “wing” section of the spectrum to the total number of counts in the peak (Fig. 10). This parameter represents a relative measure of the high-momentum annihilation events. In general, primarily valence electrons contribute to annihilation in the region of the *S* parameter because of their low momentum. On the other hand, most of the contribution to the *W* parameter originates from core electrons, which possess high momentum. For this reason, the *S* and *W* parameters are often referred to as the valence and core annihilation parameter, respectively. Both parameters can yield very useful information about voids and impurities present in a material. For example, voids have normally less high-momentum core electrons that positrons can annihilate with, thus leading to a higher *S* parameter.

The definition of the integration windows to determine the *S* and *W* parameters is sometimes somewhat ambiguous. As a general rule, however, the *S* parameter is defined in such a way that the number of annihilation events in the central window is large enough (typically  $>10^6$ ) compared to the total number of counts in the peak so that the statistical accuracy is maintained. Therefore, usually  $S \approx 0.5$ . When defining the *W* parameter, instead, the low-energy limit of the integration window is selected so that it is far away from the peak center.

Despite the common definitions described above, the *S* and *W* parameters depend on many experimental factors. The detector and sample geometry, the energy resolution, the peak calibration, and the amplifier gain can all significantly affect the

magnitude of both parameters. In particular, the relative angle between the direction of the measurement and the crystal lattice alignment is important, as the natural anisotropies present in the material may generate somewhat different spectra. The absolute values of the  $S$  and  $W$  parameters are, therefore, meaningless in themselves. Comparative measurements, for instance, with a defect-free reference sample, are needed to draw any meaningful conclusions from the measured spectra. For this reason, the spectra for the defect-containing samples are usually given as ratios to the spectrum of the corresponding defect-free sample. Similarly, the  $S$  and  $W$  parameters are normally given as normalized values to those of the reference sample ( $S_B$ ,  $W_B$ ). In doing so, the measured spectra do not depend on the aforementioned instrumental parameters, provided that all measurements are carried out with the same experimental setup. If data are collected with different measurement systems, comparison between normalized  $S$  and  $W$  parameters requires careful analysis and interpretation.

Given that the  $S$  and  $W$  parameters are inherently time-integrated quantities, their behavior resembles, up to some extent, that of the average positron lifetime  $\tau_{av}$  (see Eq. 10). One can define:

$$S = \eta_B S_B + \eta_D S_D, \quad (17)$$

$$W = \eta_B W_B + \eta_D W_D, \quad (18)$$

where  $S_D$  and  $W_D$  are the parameters determined by trapping at defects  $D$ . Therefore, as long as  $S_D$  and  $S_B$  are known, the defect concentration can be estimated from the  $(S, W)$  parameter pair. Unlike in PALS measurements, however, information about specific defects present in the material cannot be obtained from DBS spectra, as the data originates from the superposition of annihilations in various states. Nonetheless, if the DBS results are complemented with PALS measurements, such kind of defect-specific information may be inferred from correlations arising between the characteristic parameters of the two techniques. For instance, a linear correlation between the pairs of parameters  $(S, W)$ ,  $(S, \tau_{av})$ , and  $(W, \tau_{av})$  arises when samples with different concentrations of the same kind of defects are measured.

### Information from CDB Spectra

CDB spectra yield the momentum density distribution of the annihilating electron-positron pairs with high accuracy. As explained in detail in section “[Data Analysis](#),” positron annihilation with core electrons prevails in the section of the spectrum around several atomic units. Annihilation with different kinds of atoms leads to distinct high-momentum patterns, because each element is characterized by a unique core shell which retains its features even when atoms form a solid. In other words, this part of the spectrum contains information on the chemical nature of the atoms that the positrons annihilate with. Hence, CDB measurements can help identify defects such as impurities in the host lattice and investigate any elemental variations around a defect site. It is interesting to note that the spatial distribution of different atomic elements around the defects does not affect the shape of the spectrum as much as the chemical nature of the surroundings.

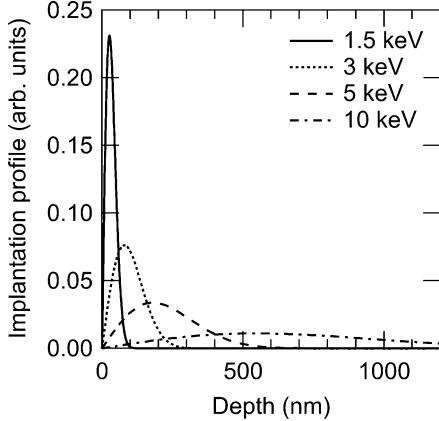
The strong elemental selectivity of the CDB method is apparent in bulk measurements of metals, semiconductors, and their alloys (Asoka-Kumar et al. 1996). In a folded, normalized spectrum, the fingerprint of a certain element may be observed as a peak at a given momentum. The intensity of the peak is correlated with the average number of nearest neighbors of that element around the vacancy defects. Measured spectra can be compared against databases of annihilation peak shapes and their characteristic features that have been compiled for an extensive number of different elements (see, e.g., Myler and Simpson 1997). Nonetheless, care needs to be taken, for example, when analyzing data for materials which contain elements with consecutive atomic numbers. As positrons may not be able to detect the subtle differences in the core electron shells, the CDB spectra for these elements might nearly coincide. Experimental data can also be directly compared to theoretical calculations that can be used to characterize materials and identify impurity-vacancy complexes (Ghosh et al. 2000). Two-particle models that describe the positrons annihilating with core electrons are typically employed in this kind of computations (Alatalo et al. 1996).

## Variable Energy Positron Annihilation Spectroscopy

As mentioned in section “[Positron Sources](#),” although the direct implantation of radioisotope-based positrons is a straightforward technique to investigate the bulk of materials, this approach is unsuitable to probe the near-surface layers, because positrons can penetrate deep into the bulk. For this reason, energy-tunable and mono-energetic positron beams have been developed. High-energy positrons (either from  $\beta^+$  decay or pair-production) can be slowed down through the use of so-called moderators which re-emit positrons with  $\sim$ eV energies. The most common moderators are semicrystalline thin films of metallic elements, such as W and Ni, or cryogenic noble gases. The crystal surface of those thin films has a negative positron work function: any thermalized positrons within the bulk that have reached the surface can be re-emitted back to the vacuum with an energy corresponding to the work function, typically a few eV. For example, the positron work functions of tungsten and nickel are  $-3.0$  eV and  $-1.0$  eV, respectively (Schultz and Lynn 1988). Those slow positrons can then be magnetically confined and electrostatically accelerated up to 50 keV to form an energy-variable and mono-energetic beam. This beam can be directed onto the sample of interest with the desired incident energy for PAS studies.

The major advantage of positron beams is the ability to control the positron implantation depth by selecting the beam energy. As the mean stopping depth ranges from a few nm to a few  $\mu\text{m}$  (see Fig. 11), low-energy positrons can be used to probe near-surface layers and thin films. This represents the crucial feature of VEPAS (Fujinami 1996). The PAS techniques described in sections “[Positron Annihilation Lifetime Spectroscopy](#)” and “[Doppler Broadening Spectroscopy](#)” can be applied when using variable-energy positron beams to investigate the surface properties, just like the bulk was studied using radioisotope positron sources.  $S$  parameter vs. energy

**Fig. 11** Positron implantation profiles in Si at various positron beam energies. The stopping profile of positrons penetrating the sample is described by the Makhovian function (Saarinen et al. 1998)



curves, CDB, DBS, and PALS spectra can all be collected using VEPAS and provide detailed information on depth profile, chemical nature, size, and concentration of defects in the near-surface layers.

By changing the incident beam energy, and therefore the mean implantation depth, VEPAS also offers the capability of depth profiling in samples. CDB spectra can be measured to analyze the shape of the 511 keV peak as a function of the impact energy. The positron energy dependency of the  $S$  parameter (hereinafter abbreviated as  $SE$  curve) is very important as it allows determination of the depth profile of defects. As described in section “Doppler Broadening Spectroscopy,” when positrons annihilate at defects, the shape of the annihilation peak becomes sharper, and the  $S$  parameter increases in magnitude compared to positrons annihilating in the bulk.

In order to determine the depth profile of the defect distribution, the positron behavior in the material must be considered. The fast positrons from the  $^{22}\text{Na}$  source typically do not annihilate before they are thermalized in the bulk. Hence, the positron implantation distribution  $P(z, E)$  as a function of depth  $z$  (nm) and positron energy  $E$  (keV) can be described by the Makhovian distribution:

$$P(z, E) = \frac{mz^{m-1}}{z_0^m} e^{-\left(\frac{z}{z_0}\right)^m}, \quad (19)$$

$$z_0 = \frac{\langle z \rangle}{\Gamma\left(1 + \frac{1}{m}\right)}, \quad (20)$$

where  $m = 2$ . The average implantation depth  $z$  is given by:

$$\langle z \rangle = \frac{A}{\rho} E^n, \quad (21)$$

where  $\rho$  is the material density ( $\text{g}/\text{cm}^3$ ) and  $A = 40$  and  $n = 1.6$  are parameters. Figure 11 shows typical positron implantation profiles in Si at various positron incident energies between 1.5 and 10 keV calculated using the Makhovian distribution.

The thermalized positrons then diffuse and become trapped at defect sites or simply annihilate in the bulk. If the electric field in the surface layer is neglected, the positron diffusion equation which describes these processes is given by:

$$\frac{\partial n(z,t)}{\partial t} = D_+ \frac{\partial^2 n(z,t)}{\partial z^2} - \lambda_{\text{eff}} n(z,t) + P(z,E), \quad (22)$$

where  $D_+$  is the diffusion coefficient and  $n(z, t)$  is the positron density at depth  $z$  and time  $t$ . If positron trapping sites are present in the surface layer with a specific trapping rate  $\mu_D$  and a defect concentration  $c_D$ , and the annihilation rate in the bulk is  $\lambda_B$ , then  $\lambda_{\text{eff}}$  can be expressed by:

$$\lambda_{\text{eff}} = \lambda_B + \mu_D c_D(z). \quad (23)$$

Since Eq. (22) actually describes a stationary state, it becomes:

$$D_+ \frac{\partial^2 n(z)}{\partial z^2} - \lambda_{\text{eff}} n(z) + P(z,E) = 0. \quad (24)$$

Ultimately, positrons annihilate at trapping sites such as defects, at surfaces (where Ps can form) or simply as delocalized states in the bulk. Hence, in general an *SE* curve can be described by the following equation:

$$S(E) = S_S f_S(E) + S_B f_B(E) + S_D f_D(E), \quad (25)$$

where  $f_S(E)$ ,  $f_B(E)$ , and  $f_D(E)$  are the fractions of positrons annihilating at the surface, in the bulk, and at the defects, respectively, and  $S_S(E)$ ,  $S_B(E)$ , and  $S_D(E)$  are the corresponding *S* parameters of those annihilation sites. The value of  $S_B$  is obtained experimentally, while  $S_S$  strongly depends on the surface condition. Assuming a rectangular or Gaussian function for the defect depth distribution, the parameters can be optimized in order to obtain the best fit to the measured *SE* curve, for example, using a least squares method. VEPFIT (van Veen et al. 1991) is an example of such kind of analysis software. Specifically, in VEPFIT the relevant shape parameters of the defect depth distribution are extracted from the experimental *SE* curve if the *S* parameter for each kind of trapping site and the diffusion coefficient in the bulk are provided as inputs.

It is important to stress that an *SE* measurement does not directly represent the defect depth distribution. Hence, it is not appropriate to discuss the defect behavior in a material, for example, simply by comparing its *SE* curve with that of the corresponding defect-free sample and taking the difference in the *S* parameter at each energy. In fact, even if the positron impact energy is the same, the region where positrons annihilate depends on the specific defect distribution.

CDB measurements in VEPAS also allow determination of the chemical state of defects. As explained in section “[Doppler Broadening Spectroscopy](#),” a change in the high-momentum component of the spectrum provides information on the elemental nature of defects around the positron trapping sites. After normalizing the annihilation peak of the sample of interest to that of a reference sample (where no annihilation in defects occurs), the ratio of the CDB curves or the correlation between the  $S$  and  $W$  parameters, i.e., a so-called ( $S, W$ ) plot, is measured as the positron energy is varied. This kind of correlation plots reflects the specific elemental species that positrons annihilate with in the vicinity of defects. This type of measurement can provide information about complex defects, such as those in pores, and bound impurities which are important in compound semiconductors.

VEPAS can also be combined with PALS measurements to obtain information about the defect size, which is not provided by CDB spectra. However, in this case the start signal for the PALS measurement is somewhat difficult to define as the positron implantation in the sample is no longer determined by the 1.27-MeV  $\gamma$ -quantum from the  $^{22}\text{Na}$  source. This issue can be overcome, for instance, by incorporating a pulsing device during the positron beam transport. In doing so, the start signal originates from the frequency generator for pulsing. The stop signal remains the annihilation  $\gamma$ -rays detected by the  $\text{BaF}_2$  scintillator. Existing pulsed VEPAS systems are both accelerator-based (Suzuki et al. 1991) and radioisotope-based (Oka et al. 2009). Other ways to measure the positron lifetime without pulsing the beam have been developed, such as a method in which the start signal is driven by the secondary electrons emitted from the sample (Lynn et al. 1984). However, in this case the time resolution is quite large, around 600 ps.

More recently, positron microbeams have also been established with the aim of measuring the 3D distribution of defects in the near-surface layers. Microbeams are produced by electrostatically transporting a positron beam and focus it to a spot with a diameter of about 10  $\mu\text{m}$  (Greif et al. 1997; Maekawa and Kawasuso 2008). Other methods include the combination of electromagnetic fields to transport and focus the beam and a remoderator for brightness enhancement (Oka et al. 2009). By scanning a sample with a positron microbeam with a variable energy, say, between 10 and 30 keV and measuring the  $S$  parameter as a function of the ( $x, y$ ) position, the 2D distribution of defects in the 1- $\mu\text{m}$ -depth region can be obtained with a spatial resolution of about 10  $\mu\text{m}$ . In addition, the spatial distribution of the average positron lifetimes at various depths can also be imaged using a positron probe microanalyzer (Oshima et al. 2011). It is to be noted, however, that the spatial resolution of the defect maps obtained this way does not just depend on the beam size but also on the time distribution of the positron pulse and the spread due to the diffusion of the thermalized positrons.

Positron beam-based technologies do not come without any disadvantages. The main shortcoming is the relatively low beam intensity compared to radioisotopes ( $^{22}\text{Na}$  yields up to  $10^9$  positrons/s). This is due to the low moderation efficiency of current moderators, typically of the order of  $10^{-3}$  to  $10^{-4}$  (Zecca 2002). Positron beam intensities are usually of the order of  $10^6$ – $10^7$ /s, although recently beams with an intensity as high as  $\sim 10^8$ /s have been achieved (see, e.g., Stanja et al. 2016). Therefore, the measurement time using beams might somewhat increase compared to radioisotope sources.

## Application Examples

In this section, we present some results on defects in various materials, such as elemental metals, semiconductors, and polymers, obtained with the PAS methods. A couple of examples for each group are presented to show how the combination of PALS, DBS, or CDB experiments and VEPAS, sample treatment, theoretical calculations, and complementary analytical techniques has been used to identify and elucidate the properties of defects in those materials. These cases have been selected in order to illustrate the different kind of features observed in the experimental data, as they offer a detailed identification of the defects. However, it is important to note that in some cases it may not be possible to systematically characterize the atomic structure of defects and their properties using the PAS techniques. This is often due to the presence of multiple kinds of defects with high enough concentration in the material or the inability to obtain/prepare a reference, i.e., the defect-free sample. Nonetheless, important information on the formation of defects and on their properties can be obtained even without a complete identification.

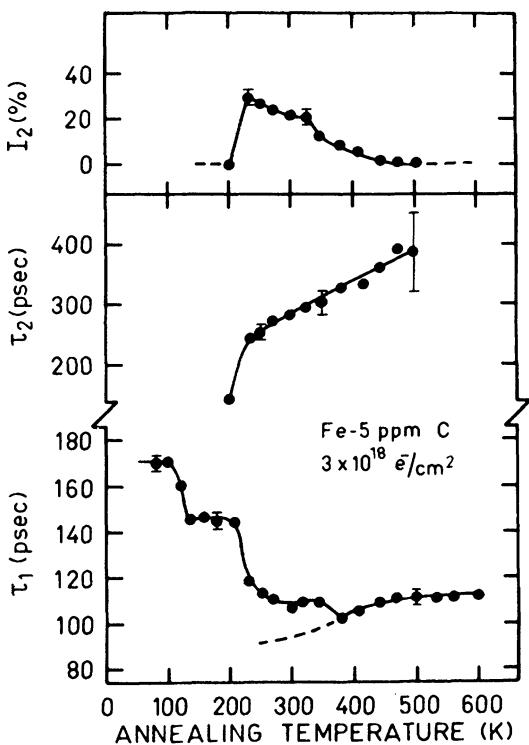
### Metals

#### Vacancy Mobility and Clustering in Electron-Irradiated and Deformed Iron

As positrons are sensitive to a large range of void sizes, PALS can be used to elucidate the dynamics of defect formation, migration, and clustering by directly inspecting the vacancy mobility as a function of temperature (Petersen et al. 1974). The understanding of these processes is particularly important to clarify the properties of metals irradiated with energetic particles or plastically deformed. These materials have important technological and industrial applications, such as the development of suitable materials for nuclear reactors. Thanks to the systematic study of a large variety of pure face-centered cubic (fcc) and body-centered cubic (bcc) metals and concentrated alloys using PALS, our understanding of this problem has dramatically improved. In particular, careful analyses of annealing kinetics revealed the temperature region of the free vacancy migration stage.

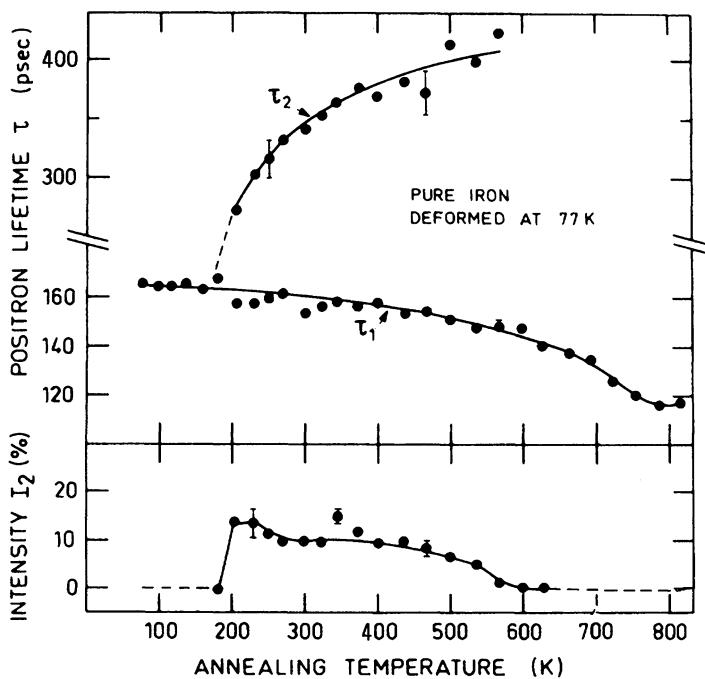
One notable example of such studies on the elementary properties of point defects is the PALS measurements of high-purity  $\alpha$ -iron by Vehanen et al. (1982). The behavior of monovacancy migration was investigated in specimens damaged by electron irradiation and plastic deformation. Figure 12 shows the PALS results as a function of the isochronal annealing temperature for an electron-irradiated sample at 20 K with a dose of  $3 \times 10^{18} \text{ e}^-/\text{cm}^2$ . Up to 100 K the spectrum consists of only one exponential component with a lifetime  $\tau_1 = 175 \text{ ps}$ . As this is the typical positron lifetime in a monovacancy for iron, this indicates that all positrons are trapped by defects, i.e., the monovacancies are generated by electron irradiation. At 120 K,  $\tau_1$  decreases because of the reduction in the monovacancy concentration due to the migration of self-interstitial atoms and the recombination of a fraction of Frenkel pairs. After annealing at 220 K, a steep decrease in  $\tau_1$  is again observed and  $\tau_1$

**Fig. 12** PALS results as a function of the isochronal annealing temperature for electron-irradiated iron with a dose of  $3 \times 10^{18} \text{ e}^-/\text{cm}^2$ . The error bars represent the statistical standard deviations of the data. The dashed line for  $\tau_1$  represents the result of the trapping model assuming the presence of vacancy clusters only. (Reprinted with permission from Vehanen et al. (1982). Copyright (1982) by the American Physical Society)



reaches the bulk value of 110 ps. At the same time, a distinct, long lifetime component  $\tau_2 = 300$  ps rises sharply with an intensity  $I_2 \approx 20\%$ . This component is ascribed to positron trapping into three-dimensional vacancy clusters. This means that monovacancies generated by the electron irradiation start to freely migrate at 220 K and form small vacancy agglomerates. Following the nomenclature in fcc metals, this temperature range of free monovacancy migration is often referred to as stage III in  $\alpha$ -iron. By further annealing above stage III,  $\tau_2$  continues to rise which indicates that the vacancy clusters keep growing in size. Finally, as  $I_2$  approaches zero at around 500 K, the vacancy agglomerates anneal out.

The PALS results for the plastically deformed iron sample are shown in Fig. 13. This specimen was subject to pressure under liquid nitrogen (77 K) to reduce its thickness by up to 60%. Below 200 K, the spectrum is composed of just one exponential component with a lifetime  $\tau_1 = 165$  ps. This indicates that all positrons are trapped at vacancies and dislocations produced by the plastic deformation. The lifetime value of  $\tau_1$  is actually a little lower than the positron lifetime in monovacancies for iron (175 ps). This result is not unusual as it is observed in many other metals subject to plastic deformation. Similar to the electron-irradiated sample, annealing in stage III leads again to the rise of a distinct, long lifetime component  $\tau_2 \approx 300$  ps with an intensity  $I_2 \approx 13\%$ . This suggests that the vacancies generated during plastic deformation also become mobile at 200 K and aggregate into small



**Fig. 13** PALS results as a function of the isochronal annealing temperature for plastically deformed iron at 77 K. The error bars represent the statistical standard deviations of the data. (Reprinted with permission from Vehanen et al. (1982). Copyright (1982) by the American Physical Society)

void clusters. As the temperature is increased further above stage III,  $\tau_2$  continues to grow, and the microvoid structures accordingly expand in size. At 500 K the vacancy clusters start to become thermally unstable and thus the intensity  $I_2$  progressively declines to zero. Above 600 K, all agglomerates anneal out and positrons are trapped at the remaining dislocations only. At the same time,  $\tau_1$  gradually decreases with temperature, which points toward a recrystallization process. At around 800 K,  $\tau_1$  finally reaches the value of 110 ps, which corresponds to the positron lifetime in a defect-free lattice (bulk). The results for the plastically deformed sample are consistent with those in electron-irradiated iron since the prominent features in the annealing process, i.e., monovacancy migration and consequent formation of small agglomerates, and vacancy cluster dissociation, are found to occur at around the same temperatures, namely, stage III (200–220 K) and 500 K, respectively.

### Embedded Copper Nanoprecipitates in Iron

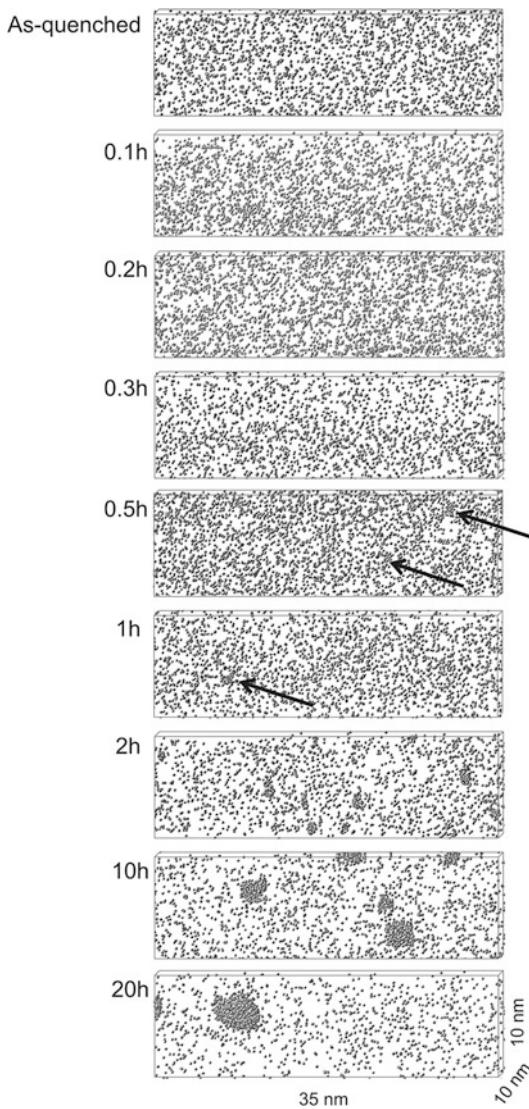
Embedded nanoparticles and subnanoparticles (ENPs) in host materials, such as ultrafine precipitates in metal alloys and quantum dots in semiconductors, play an important role in the structural and electronic properties of materials of technological

and industrial interest. The characterization of ENPs, namely, their size, concentration, chemical composition, and electronic structure, is thus crucial in order to understand the physical and mechanical properties of these materials. Although ENPs are not vacancy-type defects, positrons have been theoretically predicted to be confined in ENPs and precipitates with positron affinities lower than that of the host material (Puska et al. 1994). This is another kind of positron trapping state in a solid, called a quantum dot-like state, in which the wave function of every positron is entirely spatially confined within the ENPs, just like electrons in their quantum dots.

Nagai et al. (2000), and later Toyama et al. (2012), experimentally confirmed the positron confinement to ENPs by affinitive trapping by measuring the behavior of the embedded Cu nanoprecipitates (ECNPs) in a Fe matrix using CDB spectroscopy. ECNPs form in the early thermal aging stage of Fe-Cu alloys and play a significant role in neutron embrittlement of pressure vessel steels used in nuclear reactors. The samples (Fe-1.0 wt % Cu) were annealed at 825 °C for 4 h, quenched into iced water, and thermally aged at 475 °C for 0.1 to 20 h. In the as-quenched state, the Cu atoms are in a supersaturated solid solution and precipitate by thermal aging. Three-dimensional atom probe (3D AP) maps (Fig. 14) confirm the formation of the ECNPs. For short aging times (up to 0.3 h), no Cu atom aggregates are observed. However, after 0.5 and 1 h aging, small Cu clusters start to be observed. The presence of ECNPs is clearly confirmed after 2 h of aging as they grow in size with time.

Figure 15 shows the CDB spectra of the aged Fe-Cu specimens with different heat treatments and fcc pure Cu normalized to bcc Fe. The shape of the spectrum of the as-quenched sample is flat at around 1, which means that it is identical to that for pure bulk Fe. This means that the positrons are not trapped by the isolated Cu atoms in the supersaturated solid solution but annihilate only with the Fe atoms. As the samples start to undergo thermal aging, a peak in the high-momentum region at around 3.3 a.u. arises. Since the spectrum for pure Cu shows the same characteristic feature, this peak can be attributed to the specific momentum distribution of the closed  $3d$  shell of Cu. This suggests that some positrons become trapped at the ECNPs. It is interesting to note that in this early aging stage (up to 0.3 h), the 3D AP maps do not show any ECNPs (Fig. 14), whereas positrons can already detect the ECNPs. This proves the very high sensitivity of the positron techniques. The peak at 3.3 a.u. becomes more pronounced as the annealing progresses and after 2 h aging, the CDB spectrum is almost the same as that of pure Cu. This indicates that all positrons annihilate only with the Cu atoms and not with the Fe atoms. The concentration of the ECNPs can be estimated to be of the order of  $10^{18}/\text{cm}^3$  assuming that about 10% of Cu atoms have precipitated from the Fe matrix solid solution. Typically, for materials with a defect concentration higher than  $10^{18}/\text{cm}^3$ , all positrons are entirely trapped by the defects. If positrons were trapped at possible open-volume defects, such as vacancies or interfaces at the precipitates, their lifetime would be longer than that in pure bulk Cu, and their momentum density would be enhanced in the low-momentum region. As neither of these is experimentally observed, it can be concluded that all the positrons in the Fe-Cu samples after 2 h aging are exclusively confined into the ECNPs. Moreover, the ECNPs consist of Cu

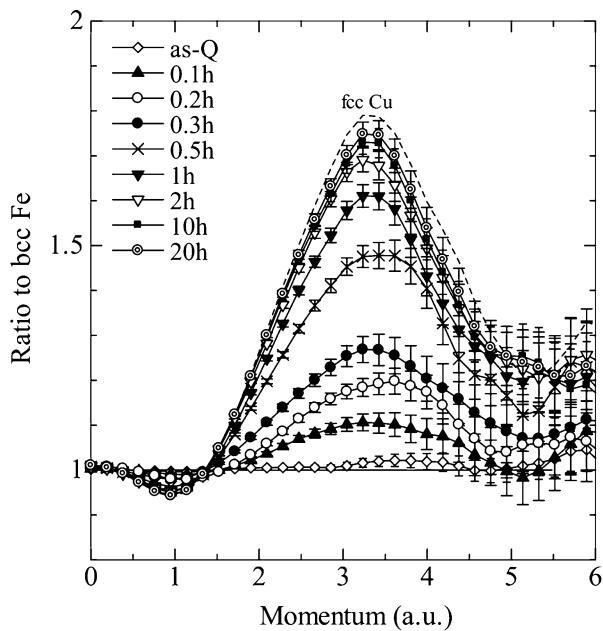
**Fig. 14** Three-dimensional atom probe maps of Cu atoms in the thermally aged Fe-Cu alloy samples. The ECNPs after 0.5 and 1 h aging are indicated by arrows.  
 (Reprinted with permission from Toyama et al. (2012). Copyright (2012) by the American Physical Society)



atoms only, are free from Fe atoms and vacancies, are three-dimensional, and possess no open-volume defects at the interfaces with the Fe matrix which can trap positrons. The fraction of positrons annihilating in the ECNPs can be estimated at each aging time from the relative intensity of the measured peak at 3.3 a.u. compared to that for pure Cu (Table 1). As expected, the fraction increases with aging time and becomes greater than ~90% after 2 h.

This quantum dot-like state of the positrons confined into the ECNPs can be explained in terms of the difference in the positron affinity between the precipitates and the matrix. Since the positron affinity of Cu ( $-4.81$  eV) is  $\sim 1$  eV lower than that

**Fig. 15** Ratio curves of the CDB spectra of the thermally aged (0 to 20 h) Fe-Cu samples to that of bcc Fe. The spectrum of fcc Cu is also given as a reference.  
 (Reprinted with permission from Toyama et al. (2012). Copyright (2012) by the American Physical Society)



**Table 1** Fraction of positrons annihilating in the ECNPs as a function of the thermal aging time. Based on the data from Toyama et al. (2012)

Aging time (h)	Fraction of positron annihilation in the ECNPs (%)
0	~0
0.1	13
0.2	24
0.3	35
0.5	61
1	76
2	87
10	92
20	95

of Fe ( $-3.84$  eV), the precipitates are viewed by the positrons as potential wells with a depth of 1 eV. Hence, if the precipitates are spherically symmetric and larger than  $\sim 0.6$  nm in diameter, positron bound states to the precipitates exist. Below this size, a positron cannot be bound because the zero-point motion energy exceeds the depth of this potential well. This means that the size of the ECNPs is about 0.6 nm after the first  $\sim 10$  min of thermal aging when the positron confinement into the ECNPs starts to be observed. The interesting aspect of the affinitive positron trapping is that it occurs whenever the positron affinity of the ENPs is lower than that of the host material. Therefore, this positron state can be used as a unique and powerful tool for the study of mesoscopic systems in a wide range of materials.

## Semiconductors

### Depth Distribution of Ion Implantation-Induced Defects in Silicon

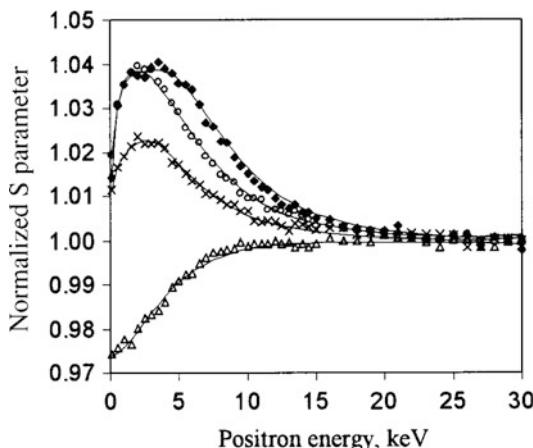
Semiconductors represent a particularly large family of technologically important materials that can be efficiently investigated using PAS. Many semiconductors are grown on thin films, and epitaxial growth is often performed on non-native substrates, thus increasing the probability of lattice mismatch and leading to highly extended defect densities (Tuomisto and Makkonen 2013). It is, therefore, of utmost importance to elucidate the nature and type of defects in these cases.

Here we give an overview of a study case of defects introduced in ion implantation processing. This is often used as an impurity introduction method in semiconductor device manufacturing, as ion implantation dramatically affects the doping efficiency. Hence, it is no surprise that the vast majority of PAS studies on vacancies in silicon have been performed on irradiated or implanted material. Fujinami et al. (2003) analyzed the depth distribution of ion implantation-induced defects in silicon from measurements of the  $SE$  curve using the VEPAS method. Prior to the measurements,  $P^+$  ions were implanted (at 100 keV and with a fluence of  $2 \times 10^{14}/\text{cm}^2$ ) into a single crystal Si wafer with an average depth of 136 nm estimated using TRIM (Ziegler et al. 1985), followed by annealing. To overcome the inherent limit in the achievable depth resolution due to broadening of the positron implantation profile, iterative chemical etching of the sample surface was performed before the measurement. This chemical polishing is believed to introduce no further defects in the material.

Figure 16 shows the measured  $SE$  curve of the unimplanted sample, as-implanted sample without etching, and after etching of a 220-nm- and 450-nm-thick surface layer. In the 220-nm etched sample, a surface layer roughly corresponding to the ion implantation region was removed, whereas in the 450-nm etched sample, a surface layer deeper than that was removed. For both these samples, however, the  $S$  parameter is larger than the bulk value, indicating that defects have been introduced deeper in the sample than those depths.

In the analysis of the  $SE$  curves using VEPFIT, it is presumed that all generated defects are divacancies  $V_2$  (i.e., a pair of vacancy defects), as they are characterized by a typical  $S$  value of 1.04, and the defect depth profile is fitted by a rectangular distribution. The defect distribution is determined by consecutively analyzing the data of the 450-nm, 220-nm etched and unetched samples using the output distribution from the previous sample as input for the next. The defect distribution obtained in this manner (Fig. 17a) shows that defects spread to a region several times deeper than the average implantation depth of the ions. Moreover, the defect distribution displays a shoulder on the surface side of the injection region. Since in TRIM the diffusion of vacancies is not modeled, the defect formation in this region is not due to so-called atomic protrusion but rather by diffusion of the voids directly ejected by the ion implantation.

The vacancy mobility and clustering are also investigated by measuring the  $SE$  curves of the samples annealed at 300 °C and 500 °C (Fig. 18). Figure 17b, c shows the corresponding depth distribution of defects obtained with VEPFIT. After



**Fig. 16** Normalized  $S$  parameter (to the  $S$  parameter of the bulk) as a function of positron energy for  $P^+$  ion-implanted Si. The  $SE$  curves are measured after chemical etching of a layer of 0-nm ( $\blacklozenge$ ), 220-nm ( $\circ$ ), and 450-nm ( $\times$ ) thickness. Solid lines represent the fit to the data modeling the defect profile shown in Fig. 17. The  $SE$  curve for the unimplanted sample ( $\triangle$ ) is also shown. (Reprinted from Fujinami et al. (2003), with the permission of AIP Publishing)

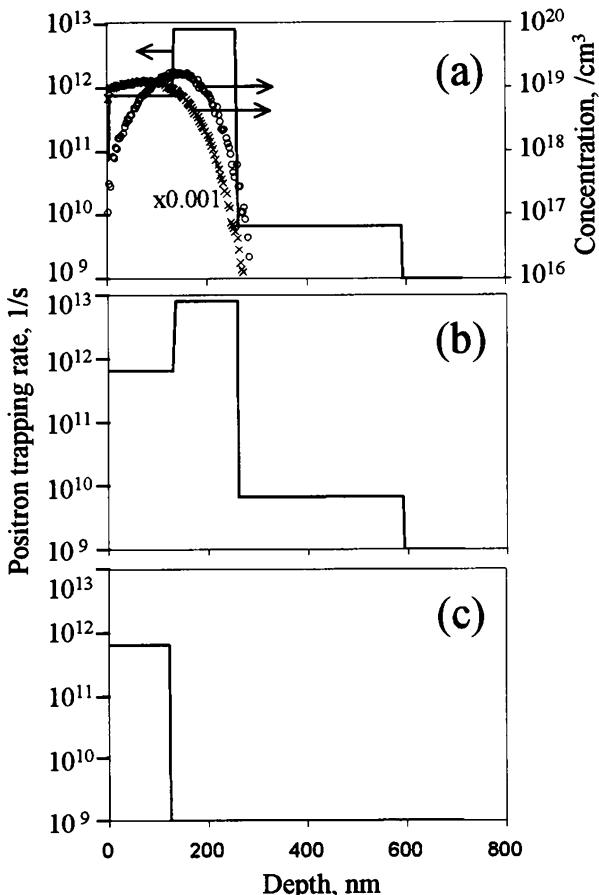
annealing at 300 °C, the  $S$  parameter increases compared to the as-implanted sample, and, therefore, a good fit can be achieved only by setting the  $S$  value for  $V_2$  to 1.05 and with almost the same defect distribution as that of the as-implanted sample. This suggests that divacancies become unstable above ~230 °C and merge into larger vacancy clusters. After annealing at 500 °C instead, the peak in the  $SE$  curve shifts to lower energies, and the defect distribution becomes shallower than the implantation region.

### EL2-Induced Defect in Gallium Arsenide

Compound semiconductors are particularly important due to their unique electrical and optical properties and several technological applications. Hence, they have been the subject of the extensive investigation of point defects and their effect on the material properties using the PAS techniques. The so-called EL2 midgap donor defect in gallium arsenide represents one of the most noticeable examples of this kind of studies. This defect is known to play a major role in the growth of undoped semi-insulating (SI) GaAs (Kaminska and Weber 1993) due to its optically induced metastability. In fact, EL2 transitions to the neutral metastable state EL2\* under photoillumination with an energy of 0.8–1.5 eV at temperatures below 100 K. PALS and DBS helped to clarify the atomic structure of this defect which was previously unknown.

Figure 19 shows the results of the PAS measurements on two SI GaAs samples grown with different EL2 concentrations after illumination with 1.2-eV photons at different temperatures. The top panel in Fig. 19 confirms that the EL2 line is photoquenched below ~100 K through complementary infrared absorption measurements. Both  $\tau_{av}$  and the  $S$  parameter are found to increase below this temperature,

**Fig. 17** TRIM calculation of the P<sup>+</sup> (○) and vacancy (×) depth profiles in P<sup>+</sup>-implanted Si. The defect profiles used to model the SE curves for (a) the implanted sample, (b) the sample annealed at 300 °C, and (c) the sample annealed at 500 °C are shown. (Reprinted from Fujinami et al. (2003), with the permission of AIP Publishing)

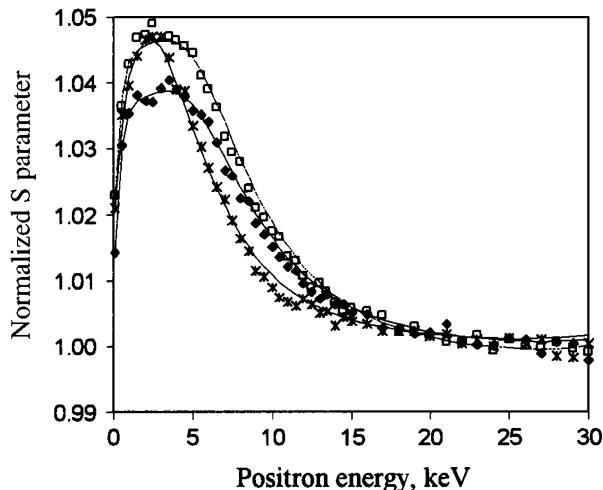


which confirms that the EL2 photoquenching produces vacancy defects. The increase in  $\tau_{av}$  with the EL2 concentration suggests a corresponding growth in the vacancy concentration (Saarinen et al. 1994). Moreover, measurements of the photon energy dependence of this effect confirm that the vacancy generation is due to the EL2 conversion to the metastable state (Saarinen et al. 1994).

The metastable vacancy defect was identified through analysis of the positron lifetime. Two kinds of vacancies are expected to trap positrons: the Ga and As vacancies created during the growth process and the metastable EL2 vacancies after photoillumination. Since the typical lifetime value of the Ga and As vacancies is 260–300 ps, whereas the measured lifetime is ~245 ps, the latter can be attributed to the formation of metastable EL2 vacancies. As this lifetime is shorter, it can also be concluded that the EL2 defect is characterized by an open volume that is smaller in size than that of a native Ga or As monovacancy.

The results from the PAS experiments validated the model for the EL2 defect, which is thought to consist of an isolated As antisite defect that relaxes toward the

**Fig. 18** SE curves for the 100 keV P<sup>+</sup>-implanted Si sample annealed at 300 °C (□), 500 °C (×), and the as-implanted sample (◆). The solid lines represent the fit to the data using the defect profiles shown in Fig. 17. (Reprinted from Fujinami et al. (2003), with the permission of AIP Publishing)



interstitial position in the metastable state (Chadi and Chang 1988; Dabrowski and Scheffler 1988). This location supports the smaller open-volume size of the EL2 defect probed by the measured positron lifetime in between the bulk and the Ga-As vacancy values.

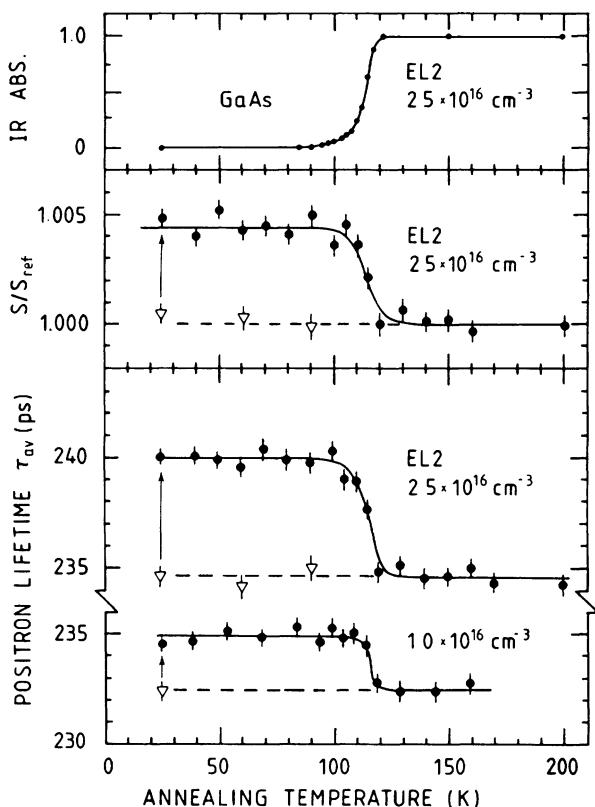
## Polymers

### Temperature Dependence of Free Volume in Amorphous Polymers

Free volume is an important property that characterizes amorphous polymers as, for instance, it plays a significant role in the volume change when polymers undergo phase transitions. In liquid states, polymer chains can move freely, and, therefore, the free volume and the overall polymer volume increase as a function of temperature. On the other hand, in glassy states the motion of polymer chains is inhibited, and, hence, the free volume is assumed to be almost constant below the glass transition temperature. In this case just the volume occupied by the substance itself changes as a function of temperature because of thermal expansion.

In order to describe the free-volume contribution to the overall volume change as a function of temperature, Hagiwara et al. (2000) investigated the behavior of some amorphous polymers around the glass transition temperature. Volume expansion measurements with the pressure-volume-temperature (PVT) technique were complemented by *in situ* Ps annihilation lifetime spectroscopy measurements. The average free-volume size was calculated from the experimentally determined o-Ps lifetime using the Tao-Eldrup equation (Eq. 14).

Figure 20 shows the measured o-Ps lifetime and the corresponding free-volume size estimated using this equation for four different amorphous polymers as a function of temperature. The free volume grows in size in the glassy state, and a discontinuity in this trend appears right at the respective glass transition temperature of each polymer. Above that temperature, the temperature dependence of the free

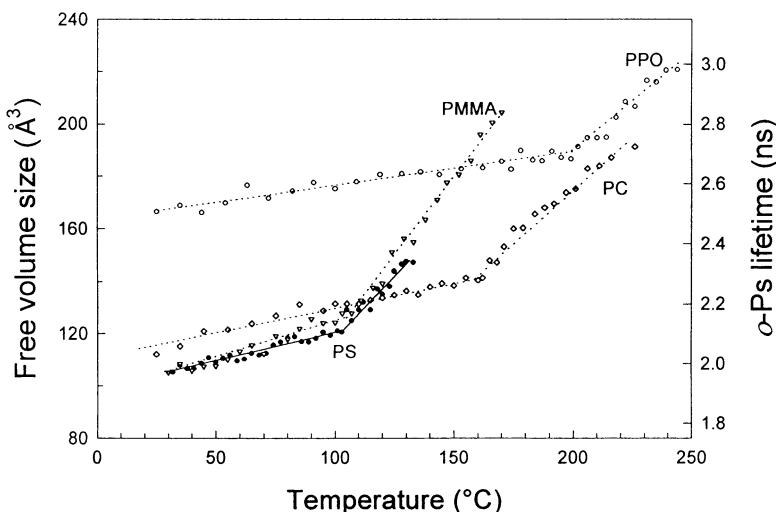


**Fig. 19** The average positron lifetime  $\tau_{av}$  and the annihilation line-shape parameter  $S$  in semi-insulating GaAs as functions of the isochronal annealing temperature after 1.2-eV illumination at 25 K. The illumination transforms EL2 into the metastable state, and corresponding changes in positron parameters are indicated by arrows. The open triangles with dashed lines below 100 K represent the reference levels where EL2 is in the stable state. The normalized infrared absorption coefficient is shown in the top panel. All measurements have been made in darkness at 25 K. (Reprinted with permission from Krause et al. (1990). Copyright (1990) by the American Physical Society)

volume increases strongly. This result shows the sensitivity of Ps to phase transitions through the consequent change in free volume. It also supports the conclusion that the free volume expands considerably and plays an important role in the overall volume change not only in the liquid state but also in the glassy state. In fact, the volume change of amorphous polymers seems to be dominated by the free volume both in the liquid and glassy states.

### Internal Structure of Reverse Osmosis Membranes

Reverse osmosis (RO) is a common water purification technology based on the separation of ions and molecules from drinking water using composite membranes. Over the last decades, developments in membrane technology and manufacturing process have significantly improved membrane permeability, RO energy efficiency,



**Fig. 20** Change of free-volume size and o-Ps lifetime in amorphous polymers under a pressure of 0.1 MPa as a function of temperature. The samples are polystyrene (PS), poly(methyl methacrylate) (PMMA), bisphenol-A polycarbonate (PC), and poly(2,6-dimethyl phenyleneoxide) (PPO). (Reprinted from Hagiwara et al. (2000), with permission from Elsevier)

and cost for a range of applications (Lee et al. 2011). However, in order to optimize the RO process, a better knowledge of the internal structure of those thin films must be gained. In particular, the molecular transport mechanism of permeates through the molecular-level holes in the separation active layer of the membranes must be fully understood. For this reason, various well-known techniques such as scanning electron microscopy, transmission electron microscopy, and atomic force microscopy have been employed to try to characterize the hole structure of the thin active layer. However, these techniques are inadequate to determine free-volume holes. This is why, in recent years, Ps annihilation-based techniques have become the tool of choice for the systematic examination of this kind of systems (see Singh 2016 and references therein).

Here an overview is given of a recent example probing the nanostructure of RO membranes using Ps annihilation. Chen et al. (2011) carried out PALS and DBS measurements by VEPAS on the active layer of commercially available composite membranes. Rejection rates of different solute compounds were determined by a conventional pressure-driven process using aqueous solutions. Figure 21a shows the o-Ps lifetime and the corresponding free-volume radius estimated using the Tao-Eldrup equation (Eq. 14), while Fig. 21b shows the intensity ( $I_3$ ) of that component for three different membranes as a function of the positron incident energy or mean implantation depth. The three membranes show quite different o-Ps lifetime and intensity behaviors, confirming the different polymer structures of their active layers. Eventually they all approach the substrate value (~1.8 ns and ~22%) at the highest incident energy.

**Fig. 21** Variations in (a) the o-Ps lifetime and (b) the intensity of that component for the composite membranes as a function of positron incident energy or mean implantation depth. The free-volume hole radius estimated using the Tao-Eldrup equation is shown on the right-hand axis of (a). (Reprinted with permission from Chen et al. (2011). Copyright (2011) American Chemical Society)

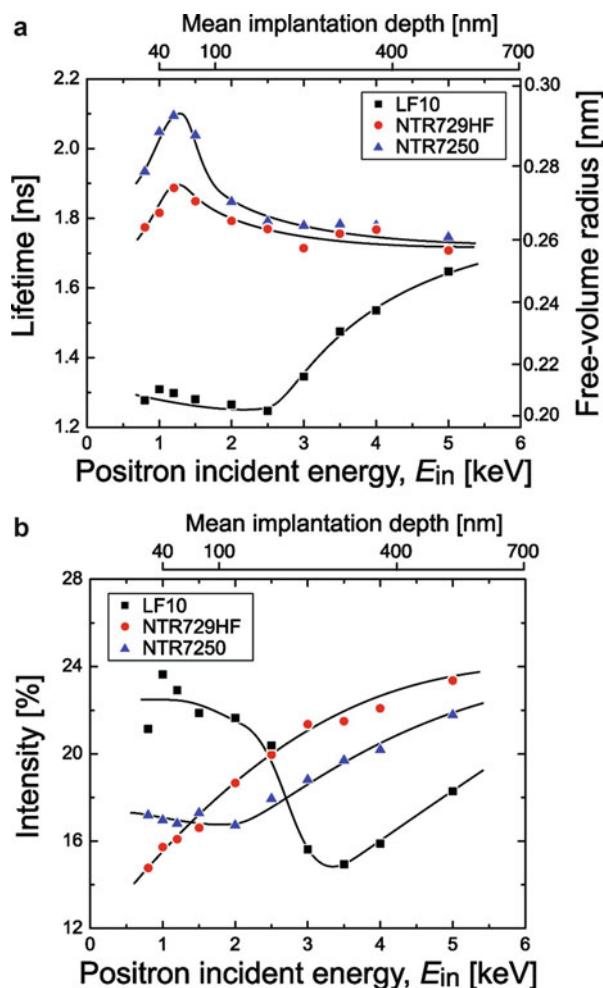
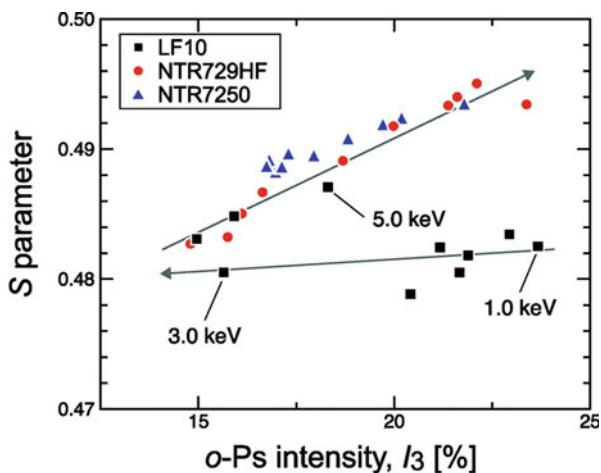
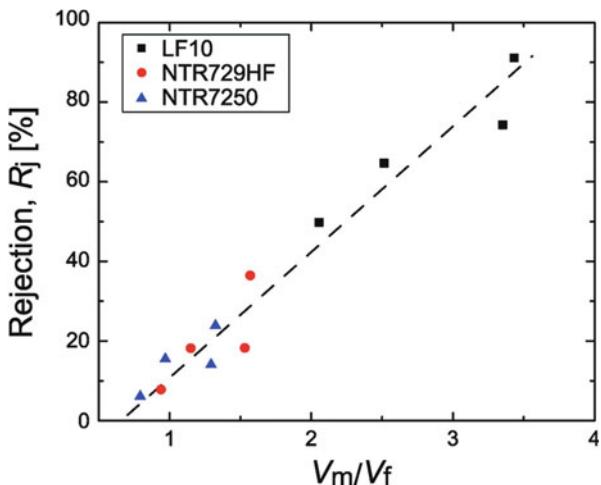


Figure 22 shows the plots of the  $S$  parameter vs o-Ps intensity measured at various incident energies ranging from 0.8 to 5.0 keV. The correlation between these two parameters is linear for two of the membranes (NTR729HF and NTR7250). This indicates that the  $S$  parameter for o-Ps is constant for both membranes as a function of depth. The smooth dependence of  $S$  on  $I_3$  for NTR729HF and NTR7250 stems from the gradual rise in the mean implantation depth with increasing positron energy. However, for the third membrane (LF10), the  $S$  parameter first decreases with decreasing o-Ps intensity up to 3 keV, and at higher energies it turns around and follows the same correlation as that for the other two membranes. The correlation between the  $S$  parameter and the o-Ps formation probability is known to be strongly dependent on the chemical structure of the polymer. Smaller  $S$  values for LF10 at low positron energies suggest a different chemical structure of the top layer of LF10 from the other two membranes. This means that the active layer of LF10 consists of two different polymers, unlike the other two membranes.

**Fig. 22** Correlation between S parameter and o-Ps intensity for the composite membranes measured at different incident energies from 0.8 to 5.0 keV. The arrows denote the directions of increasing energies for LF10. (Reprinted with permission from Chen et al. (2011). Copyright (2011) American Chemical Society)



**Fig. 23** Correlation between the rejection rate of the organic compounds and the ratio of the solute size  $V_m$  to the free-volume hole size  $V_f$  for the active layer in the three membranes. (Reprinted with permission from Chen et al. (2011). Copyright (2011) American Chemical Society)



In separation membranes, if the holes in the active layer are smaller than the molecular size of a target solute, it is expected that the transport of that solute is effectively hindered. In other words, the holes in the separation membrane act as filtering channels. By comparing the rejections  $R_j$  of different organic solutes to the ratio of the solute size  $V_m$  to the hole size  $V_f$  (see Fig. 23) obtained from the o-Ps lifetime measurements, it is possible to examine the molecular transport in the various membranes. To estimate the free-volume hole radii in the active top layer of the three membranes, only the o-Ps lifetime measured at energies up to 2.2 keV was considered. As all the data in Fig. 23 lie on a line, the molecular transport is highly dependent on the size ratio of the molecules to the free-volume hole. The latter must be comparable in size or somewhat smaller than the solute molecules.

This proves the hindering effect of molecular transport by the free-volume holes in the separation active layer. Moreover, the higher rejection rates for LF10 indicate the presence of smaller holes in that membrane. These results confirm that the Ps annihilation techniques enable the examination of the hole structure, as well as the layer structure, of composite membranes.

---

## Conclusions and Future Perspectives

Over the last decades, positron annihilation spectroscopy has established itself as one of the most commonly adopted, nondestructive methods to characterize the atomic structure of a large variety of materials. PAS takes advantage of the unique properties of antimatter to study its interaction with matter and examine the microscopic conformation of materials. The PAS methods have significantly boosted defect spectroscopy in solids enabling the identification of vacancy-type defects and their nature, size, and concentration with a high level of detail. This opened up a new avenue of materials research characterized by studies on the formation of defects, their migration process, defect clustering, role of defects in doping of semiconductors, determination of defect charge states, defect transitions, free volume in polymers, and analysis of hole structure in membranes, to cite a few among a long list of applications. In conjunction with theoretical models, all these PAS studies allowed the effective examination of the role of defects and their effect on material properties. This information is essential to design and develop materials with superior performance, advanced materials for technological application and new materials.

The various methods described here and the examples for various materials illustrated in this chapter clearly show the PAS strengths in advanced defect analysis. Nonetheless, there are still important developments that need to be achieved in order to fulfill its full potential in materials research. Here, an overview of a selection of those potential areas of progress is briefly presented. By no means this is meant to be an exhaustive listing of all possible developments in the field. Future challenges include both improvements in the experimental methods and developments in quantitative theoretical models. From a theoretical point of view, modeling of noncrystalline, amorphous, and molecular systems is very challenging, and the interpretation of data from the analysis of defect-related phenomena in these systems is relatively in its infancy. Although in recent years experimental studies on this family of materials have become abundant (Yu 2011 and references therein), a complete description of positron states, thermalization, and trapping in the free volume of these systems is still lacking. A further area where theoretical support is particularly sought is represented by studies on surfaces and interfaces between crystalline solids where positron can be trapped in so-called surface states (Nieminen and Manninen 1974). Numerous experimental investigations on these systems have been reported, but their results need detailed interpretation (Lynn and Kong 1992).

From an experimental perspective, some of the materials where current, and future PAS studies seem particularly promising are nanoparticles (Kumar et al.

2009), complex oxides (Guagliardo et al. 2012), membranes (Fujioka et al. 2015), carbon nanotubes (Guo et al. 2009), radiation-resistant alloys (Kršjak et al. 2008), soft matter, and self-assembling biomimetic systems (Fong et al. 2015). Given the complicated nature of these systems, PAS applications are often challenging and defect identification is rather complicated. Improvements in the experimental methods and interpretation of the data are needed in order to shed light onto these materials and to achieve advanced sample state manipulations. In addition, developments in the positron annihilation experimental methods themselves are also warranted. The definition of the start signal in PALS measurements using slow-positron beams is a matter of debate, because the moderation process strongly limits the usefulness of this start signal. The modulation of the positron beam might be an approach to solve this issue (Reurings and Laakso 2007). The relatively low slow-positron beam intensity is also often regarded as a major issue. This issue can be addressed by improving the current positron moderation efficiency techniques in laboratory-based facilities and by further increasing the positron source intensity in large-scale facilities, such as particle accelerators and nuclear reactors. Future improvements in moderator efficiencies would have important consequences, such as a significant reduction of measurement times, access to an isotope-source-based slow-positron beam with reasonable intensity, and the use of lower activity sources which yield less background and noise in the experiments.

---

## References

- Ackermann U, Egger W, Sperr P, Dollinger G (2015) Time- and energy-resolution measurements of BaF<sub>2</sub>, BC-418, LYSO and CeBr<sub>3</sub> scintillators. Nucl Instrum Methods Phys Res A 768:5–11
- Alatalo M, Barbiellini B, Hakala M, Kauppinen H, Korhonen T, Puska MJ, Saarinen K, Hautojärvi P, Nieminen RM (1996) Theoretical and experimental study of positron annihilation with core electrons in solids. Phys Rev B 54:2397–2409
- Anderson CD (1933) The positive electron. Phys Rev 43:491–494
- Asoka-Kumar P, Alatalo M, Ghosh VJ, Kruseman AC, Nielsen B, Lynn KG (1996) Increased elemental specificity of positron annihilation spectra. Phys Rev Lett 77:2097–2100
- Bečvář F, Čížek J, Procházka I (2008) High-resolution positron lifetime measurement using ultra fast digitizers Acqiris DC211. Appl Surf Sci 255:111–114
- Bell RE, Graham RL (1953) Time distribution of positron annihilation in liquids and solids. Phys Rev 90:644–654
- Beringer R, Montgomery CG (1942) The angular distribution of positron annihilation radiation. Phys Rev 61:222–224
- Bertolaccini L, Zappa S (1967) Source-supporting foil effect on the shape of positron time annihilation spectra. Nuovo Cimento B 52:487–494
- Bertolaccini M, Bisi A, Gambarini G, Zappa L (1971) Positron states in ionic media. J Phys C 4:734
- Brandt W (1974) Positron dynamics in solids. Appl Phys 5:1–23
- Brandt W, Paulin R (1977) Positron implantation-profile effects in solids. Phys Rev B 15:2511–2518
- Chadi DJ, Chang KJ (1988) Metastability of the isolated arsenic-antisite defect in GaAs. Phys Rev Lett 60:2187–2190
- Charlton M, Humberston JW (eds) (2001) Positron physics. Cambridge University Press, New York
- Chen Z, Ito K, Yanagishita H, Oshima N, Suzuki R, Kobayashi Y (2011) Correlation study between free-volume holes and molecular separations of composite membranes for reverse osmosis

- processes by means of variable-energy positron annihilation techniques. *J Phys Chem C* 115:18055–18060
- Coleman PG (ed) (2000) Positron beams and their applications. World Scientific, Singapore
- Connors DC, West RN (1969) Positron annihilation and defects in metals. *Phys Lett A* 30:24–25
- Dabrowski J, Scheffler M (1988) Theoretical evidence for an optically inducible structural transition of the isolated As antisite in GaAs: identification and explanation of EL2? *Phys Rev Lett* 60:2183–2186
- Dirac PAM (1928) The quantum theory of the electron. *Proc R Soc A* 117:610–624
- Eldrup M, Lightbody D, Sherwood NJ (1981) The temperature dependence of positron lifetimes in solid pivalic acid. *Chem Phys* 63:51–58
- Ferrell RA (1956) Theory of positron annihilation in solids. *Rev Mod Phys* 28:308–337
- Fong C, Dong AW, Hill AJ, Boyd BJ, Drummond CJ (2015) Positron annihilation lifetime spectroscopy (PALS): a probe for molecular organisation in self-assembled biomimetic systems. *Phys Chem Chem Phys* 17:17527
- Fujinami M (1996) Oxygen-related defects in Si studied by variable-energy positron annihilation spectroscopy. *Phys Rev B* 53:13047–13050
- Fujinami M, Miyagoe T, Sawada T, Akahane T (2003) Improved depth profiling with slow positrons of ion implantation-induced damage in silicon. *J Appl Phys* 94:4382–4388
- Fujioka T, Oshima N, Suzuki R, Price WE, Nghiem LD (2015) Probing the internal structure of reverse osmosis membranes by positron annihilation spectroscopy: gaining more insight into the transport of water and small solutes. *J Membr Sci* 486:106–118
- Ghosh VJ, Alatalo M, Asoka-Kumar P, Nielsen B, Lynn KG, Kruseman AC, Mijnarens PE (2000) Calculation of the Doppler broadening of the electron-positron annihilation radiation in defect-free bulk materials. *Phys Rev B* 61:10092–10099
- Gidley DW, Peng HG, Vallery RS (2006) Positron annihilation as a method to characterize porous materials. *Annu Rev Mater Res* 36:49–79
- Greif H, Haaks M, Holzwarth U, Männig U, Tongbhoyai M, Wider T, Maier K, Bähr J, Huber B (1997) High resolution positron-annihilation spectroscopy with a new positron microprobe. *Appl Phys Lett* 71:2115
- Guagliardo PR, Vance ER, Zhang Z, Davis J, Williams JF, Samarin SN (2012) Positron annihilation lifetime studies of Nb-doped TiO<sub>2</sub>, SnO<sub>2</sub>, and ZrO<sub>2</sub>. *J Am Ceram Soc* 95:1727
- Guo WF, Chen XL, Du HJ, Weng HM, Ye BJ (2009) Positron annihilation in carbon nanotubes. In: Wang SJ, Chen ZQ, Wang B, Jean YC (eds) Materials science forum. Trans Tech Publications, Churerstrasse, pp 198–200
- Hagiwara K, Ougizawa T, Inoue T, Hirata K, Kobayashi Y (2000) Studies on the free volume and the volume expansion behavior of amorphous polymers. *Radiat Phys Chem* 58:525–530
- Hautojärvi P (ed) (1979) Positrons in solids. Springer, Berlin
- Hodges CH (1970) Trapping of positrons at vacancies in metals. *Phys Rev Lett* 25:284–287
- Jean YC, Mallon PE, Schrader DM (eds) (2003) Principles and applications of positron & positronium chemistry. World Scientific, Singapore
- Jean YC, Van Horn JD, Hung WS, Lee KR (2013) Perspective of positron annihilation spectroscopy in polymers. *Macromolecules* 46:7133–7145
- Kaminska M, Weber R (1993) EL2 defect in GaAs. In: Weber ER (ed) Semiconductors and semimetals, vol 38. Academic, New York, pp 59–89
- Kansy J (1996) Microcomputer program for analysis of positron annihilation lifetime spectra. *Nucl Instrum Methods Phys Res A* 374:235–244
- Kirkegaard P, Eldrup M (1972) POSITRONFIT: a versatile program for analysing positron lifetime spectra. *Comput Phys Commun* 3:240–255
- Kirkegaard P, Eldrup M (1974) Positronfit extended: a new version of a program for analysing position lifetime spectra. *Comput Phys Commun* 7:401–409
- Knoll GF (2000) Radiation detection and measurement. Wiley, New York
- Krause R, Saarinen K, Hautojärvi P, Polity A, Gärtner G, Corbel C (1990) Observation of a monovacancy in the metastable state of the EL2 defect in GaAs by positron annihilation. *Phys Rev Lett* 65:3329–3332

- Krause-Rehberg R, Leipner HS (1997) Determination of absolute vacancy concentrations in semiconductors by means of positron annihilation. *Appl Phys A Mater Sci Process* 64:457
- Krause-Rehberg R, Leipner HS (eds) (1999) Positron annihilation in semiconductors: defect studies. Springer, Heidelberg
- Kršjak V, Slugen V, Mikloš M, Petřík M, Ballo P (2008) Application of positron annihilation spectroscopy on the ion implantation damaged Fe–Cr alloys. *Appl Surf Sci* 255:153–156
- Kumar N, Sanyal D, Sundaresan A (2009) Defect induced ferromagnetism in MgO nano-particles studied by optical and positron annihilation spectroscopy. *Chem Phys Lett* 477:360–364
- Lahtinen J, Vehanen A, Huomo H, Mäkinen J, Huttunen P, Rytölä K, Bentzon M, Hautajärvi P (1986) High-intensity variable-energy positron beam for surface and near-surface studies. *Nucl Instrum Methods Phys Res B* 17:73–80
- Lee KP, Arnot TC, Mattia D (2011) A review of reverse osmosis membrane materials for desalination – development to date and future potential. *J Membr Sci* 370:1–22
- Lynn KG, Kong Y (1992) Positron surface states. *Solid State Phenom* 28–29:275–292
- Lynn KG, Frieze WE, Schultz PJ (1984) Measurement of the positron surface-state lifetime for Al. *Phys Rev Lett* 52:1137–1140
- Maekawa M, Kawasuso A (2008) Construction of a positron microbeam in JAEA. *Appl Surf Sci* 255:39–41
- Makkonen I, Puska MJ (2007) Energetics of positron states trapped at vacancies in solids. *Phys Rev B* 76:054119
- Mallon PE, Schrader DM (eds) (2003) Principles and applications of positron & positronium chemistry. World Scientific, Singapore
- Manninen M, Nieminen RM (1981) Positron detrapping from defects: a thermodynamic approach. *Appl Phys A Mater Sci Process* 26:93–100
- Mogensen OE (ed) (1995) Positron annihilation in chemistry. Springer, Berlin
- Myler U, Simpson PJ (1997) Survey of elemental specificity in positron annihilation peak shapes. *Phys Rev B* 56:14303–14309
- Nagai Y, Hasegawa M, Tang Z, Hempel A, Yubuta K, Shimamura T, Kawazoe Y, Kawai A, Kano F (2000) Positron confinement in ultrafine embedded particles: quantum-dot-like state in an Fe–Cu alloy. *Phys Rev B* 61:6574–6578
- Nakanishi H, Wang SJ, Jean YC (1988) Microscopic surface tension studied by positron annihilation. In: Sharma SC (ed) Positron annihilation studies of fluids. Word Scientific, Singapore, pp 292–298
- Nieminen R, Manninen M (1974) Positron surface states in metals. *Solid State Commun* 15:403–406
- Ohkubo H, Tang Z, Nagai Y, Hasegawa M, Tawara T, Kiritani M (2003) Positron annihilation study of vacancy-type defects in high-speed deformed Ni, Cu and Fe. *Mater Sci Eng A* 350:95–101
- Oka T, Jinno S, Fujinami M (2009) Analytical methods using a positron microprobe. *Anal Sci* 25:837–844
- Olsen JV, Kirkegaard P, Pedersen NJ, Eldrup M (2007) PALSfit: a new program for the evaluation of positron lifetime spectra. *Phys Status Solidi C* 4:4004–4006
- Oshima N, Suzuki R, Ohdaira T, Kinomura A, Kubota S, Watanabe H, Tenjinbayashi K, Uedono A, Fujinami M (2011) Imaging of the distribution of average positron lifetimes by using a positron probe microanalyzer. *J Phys Conf Ser* 262:012044
- Petersen K, Thrane N, Cotterill RMJ (1974) A positron annihilation study of the annealing of, and void formation in, neutron-irradiated molybdenum. *Philos Mag* 29:9–23
- Puska MJ, Nieminen RM (1994) Theory of positrons in solids and on solid surfaces. *Rev Mod Phys* 66:841–897
- Puska MJ, Corbel C, Nieminen RM (1990) Positron trapping in semiconductors. *Phys Rev B* 41:9980–9993
- Puska MJ, Šob M, Brauer G, Korhonen T (1994) First-principles calculation of positron lifetimes and affinities in perfect and imperfect transition-metal carbides and nitrides. *Phys Rev B* 49:10947–10957

- Reurings F, Laakso A (2007) Analysis of the time resolution of a pulsed positron beam. *Phys Status Solidi C* 4:3965–3968
- Saarinen K, Kuisma S, Hautojärvi P, Corbel C, LeBerre C (1994) Metastable vacancy in the *EL2* defect in GaAs studied by positron-annihilation spectroscopies. *Phys Rev B* 49:8005–8016
- Saarinen K, Hautojärvi P, Corbel C (1998) Positron annihilation spectroscopy of defects in semiconductors. In: Stavola M (ed) Identification of defects in semiconductors, vol 51A. Academic, San Diego, pp 209–285
- Saito H, Nagashima Y, Kurihara T, Hyodo T (2002) A new positron lifetime spectrometer using a fast digital oscilloscope and BaF<sub>2</sub> scintillators. *Nucl Instrum Methods Phys Res A* 487:612–617
- Schrader DM, Jean YC (1983) Positron and positronium chemistry. Elsevier, Amsterdam
- Schultz PJ (1988) A variable-energy positron beam for low to medium energy research. *Nucl Instrum Methods Phys Res B* 30:94–104
- Schultz PJ, Lynn KG (1988) Interaction of positron beams with surfaces, thin films, and interfaces. *Rev Mod Phys* 60:701–779
- Seeger A (1974) The study of defects in crystals by positron annihilation. *Appl Phys* 4:183–199
- Siegel RW (1980) Positron-annihilation spectroscopy. *Annu Rev Mater Sci* 10:393–425
- Singh AN (2016) Positron annihilation spectroscopy in tomorrow's material defect studies. *Appl Spectrosc Rev* 51:359–378
- Stanja J, Hergenhahn U, Niemann H, Paschkowski N, Sunn Pedersen T, Saitoh H, Stenson EV, Stoneking MR, Hugenschmidt C, Piochacz C (2016) Characterization of the NEPOMUC primary and remoderated positron beams at different energies. *Nucl Instrum Methods Phys Res A* 827:52–62
- Stewart AT (1957) Momentum distribution of metallic electrons by positron annihilation. *Can J Phys* 35:168–183
- Suzuki R, Kobayashi Y, Mikado T, Ohgaki H, Chiwaki M, Yamazaki T, Tomimasu T (1991) Slow positron pulsing system for variable energy positron lifetime spectroscopy. *Jpn J Appl Phys* 30: L532–L534
- Tao SJ (1972) Positronium annihilation in molecular substances. *J Chem Phys* 56:5499–5510
- Toyama T, Tang Z, Inoue K, Chiba T, Ohkubo T, Hono K, Nagai Y, Hasegawa M (2012) Size estimation of embedded Cu nanoprecipitates in Fe by using affinitively trapped positrons. *Phys Rev B* 86:104106
- Tuomisto F, Makkonen I (2013) Defect identification in semiconductors with positron annihilation: experiment and theory. *Rev Modern Phys* 85:1583–1631
- van Veen A, Schut H, de Vries J, Hakvoort RA, Ijpma MR (1991) Analysis of positron profiling data by means of “VEPFIT”. *AIP Conf Proc* 218:171–198
- Vehanen A, Hautojärvi P, Johansson J, Yli-Kauppila J, Moser P (1982) Vacancies and carbon impurities in  $\alpha$ -iron: electron irradiation. *Phys Rev B* 25:762–780
- West R (1979) Positron studies of lattice defects in metals. In: Hautojärvi P (ed) *Positrons in solids*. Springer, Berlin, pp 89–144
- Yu Y (2011) Positron annihilation lifetime spectroscopy studies of amorphous and crystalline molecular materials. Dissertation, Martin-Luther-Universität Halle-Wittenberg
- Zecca A (2002) Positron beam development and design. *Appl Surf Sci* 194:4–12
- Ziegler JF, Biersack JP, Littmark U (1985) The stopping and range of ions in solids. Pergamon, New York

---

## **Part VII**

### **Thermal Techniques**



# Physical Basics of Thermal Techniques of Nondestructive Evaluation

36

Vladimir P. Vavilov

## Contents

Introduction .....	1350
ITT Terminology .....	1350
Basic Inspection Procedure .....	1351
Basics of Heat Transfer .....	1357
Heat Conduction, Convection, and Radiation .....	1357
Boundary Conditions .....	1359
Heat Transfer in Defects .....	1360
Differential Equation of Heat Conduction .....	1361
Analytical Solutions to ITT Problems .....	1362
Basic Relationships in ITT .....	1362
Comparing Test Procedures and Processing Algorithms .....	1364
Data Processing Algorithms .....	1368
References .....	1368

## Abstract

Thermal techniques of nondestructive evaluation (NDE) have become mature high-technology fields that have essentially contributed to nondestructive testing, condition monitoring, and predictive maintenance. In many industrial applications, these techniques help to identify problems that must be addressed, prevent equipment failure, and provide significant cost savings.

Thermal NDE has many advantages including a wide range of applications and a much higher inspection rate than other NDE methods. Improvements in IR imagers in the last decade, accompanied by dramatic advances in computer data processing, have resulted in significant improvements in the detection and

---

V. P. Vavilov (✉)  
Tomsk Polytechnic University, Tomsk, Russia  
e-mail: [vavilov@tpu.ru](mailto:vavilov@tpu.ru)

characterization of hidden structural defects in solids by analyzing material surface temperature distributions.

This chapter establishes the physical basics of thermal NDE including elements of heat conduction in solids with defects and basics of thermal radiation, as well as inspection terminology, procedures, and data processing. Following ISO 10878, this inspection technique is called Infrared Thermographic Testing (ITT). In the chapter, the elements of thermal NDE are discussed at a level which is necessary for understanding the physical principles of this testing method.

---

## Introduction

Infrared (IR) thermographic testing (ITT) is a method of inspection of materials and products by using IR thermography which, in turn, is the imaging of objects by sensing their IR (thermal) radiation. In a passive procedure, one uses an IR camera (imager) to monitor object temperature distributions and take decision on object quality by temperature anomalies. Active testing procedures require additional thermal stimulation which can be based on various physical principles, such as optical radiation, eddy currents, mechanical waves, etc. The remote, fast, and safe operation of IR cameras makes this method of nondestructive evaluation (NDE) very attractive for practitioners. Moreover, unlike many traditional NDE techniques, thermal testing is applicable to both metals and nonmetals, although its efficiency strongly depends on material thermal properties and surface conditions. Therefore, decision making on object quality in ITT represents a typical problem of distinguishing between a sought signal and a noisy background.

In condition monitoring and predictive maintenance, ITT has been appreciated for many years as a handy practical tool of indisputable economic benefit. But in the inspection of layered and composite materials, thermal NDE (TNDE) is considered as a screening, rather than independent, method. In the future, significance of this technique can be enhanced by combining ITT with other NDE methods, in particular, ultrasonic and holographic.

Table 1 shows in short some ITT historical milestones thus reflecting the author's subjective outlook at this technique that has proven its validity in technical diagnostics and NDE (aspects of IR technology chronology are deliberately omitted), see Vollmer and Möllmann (2010); Vavilov and Burleigh (2015) for details.

---

## ITT Terminology

Many old and new ITT terms and concepts are described in Table 2. Most of them are related to the temperature signal  $\Delta T$  that is typically determined as a difference between an extremum temperature in a defect area and a mean temperature in the surrounding defect-free area (Fig. 1). Also, TNDE schemes and processing techniques can be compared by the dimensionless temperature contrast

**Table 1** IR technology and TNDE in short: historical milestones. (Adapted from Vavilov 2007, 2014)<sup>a</sup>

<b>1935.</b> Nichols: IR radiometric analysis of hot rolled metals (Nichols 1935).
<b>1937.</b> Vernotte: Determination of material thermal properties by using transient thermal processes (Vernotte 1937).
<b>1948.</b> Parker and Marshall: Analysis of temperature distribution in brake shoes (Parker and Marshall 1948).
<b>1949.</b> Gorill: Inspection of soldered seams on a tin can (Gorill 1949). Leslie and Wait: power transmission line surveys (Leslie and Wait 1949).
<b>1965.</b> Beller: Active TNDE of Polaris rockets (Beller 1965). Green: Active TNDE of nuclear fuel elements (Green 1965).
<b>1970s.</b> Balageas, Karpelson <sup>b</sup> , MacLaughlin, Popov <sup>b</sup> , Vavilov et al.: 1D, 2D and 3D TNDE models. Use of heat transfer classical solutions.
<b>1980s.</b> Almond, Berardi, Busse, Carlomagno, Cielo, Favro, Hartikainen, Luukkala, Mandelis, Milne, Reynolds, Rosencwaig et al.: Thermal wave theory, pulsed TNDE, thermal property measurement. Degiovanni, Maillet et al. introduced “thermal quadrupoles.”
<b>1990s.</b> Almond, Balageas, Bison, Bremond, Budadin <sup>b</sup> , Burleigh, Busse, Cramer, Degiovanni, Grinzato, Krapez, Luong, MacLachlan Spicer, Maldague, Mandelis, Marinetti, Rantala, Ringermacher, Vavilov, Walther, Winfree et al.: Practical implementation of TNDE. Development of defect characterization algorithms. IR computerized systems. Study of fatigue phenomena. Various industrial applications. IR thermographic diffusivity and effusivity measurement.
<b>2000s.</b> Avdelidis, Balageas, Batsale, Bendada, Breitenstein, Burleigh, Busse, Cernuschi, Cramer, Dillenz, Maillet, Maldague, Mandelis, Meola, Mulaveesala, Nowakowski, Oswald-Tranta, Pickering, Sakagami, Shepard, Tuli, Xiaoyan Han, Więcek, Wiggenhauser, Winfree, Wu, Zweschaper, et al.: Commercial TNDE equipment. Novel stimulation techniques. Sophisticated data processing algorithms.
<b>2010s.</b> Batsale, Gerlach, Ibarra-Castanedo, Ishikawa, Maierhofer, Maldague, Mayr, Mulaveesala, Muzaffar, Netzelmann, Omar, Oswald-Tranta, Pradere, Rajic, Safai, Salerno, Schlichting, Sfarrà, Świderski, Xiaoyan Han, Xingwang Guo, Zalameda, Vavilov, Venkatraman et al.: Eddy current IR thermography. Inductive and ultrasonic IR thermography. Portable equipment for active TNDE, particularly, in the aerospace. Defect thermal characterization.

<sup>a</sup>The concept of this table and some references have been borrowed from (Vavilov 2007, 2014). Some milestone studies are also presented by the references (Busse 1985; Almond and Patel 1996; Vavilov and Taylor 1982; Mandelis 2001; Balageas et al. 1986; Maillet et al. 2000; Luong 1992; Shepard 2001; Cramer et al. 2006; Maldague and Marinetti 1996; Breitenstein et al. 2010; Pettersson and Bengt 1980; Rajic 2002). Many top researchers who contributed to passive IR thermographic inspection are not mentioned. The emphasis is on active TNDE

<sup>b</sup>Earlier Russian publications, not easily available to Western readers (Popov et al. 1976; Kush et al. 1988)

$C = \Delta T/T$ , or running temperature contrast, defined by the sample’s excess temperature  $T$  (Fig. 1). The latter parameter is independent of heating power  $Q$ .

## Basic Inspection Procedure

In a passive mode, test objects are characterized by temperature distributions that appear naturally due to the functioning of objects or because of other technical reasons, while active inspection requires the application of external thermal

**Table 2** ITT terms

<b>Absolute measurements</b>	made with a direct reference in contrast to differential measurements.
<b>Active IR thermography (active TNDE)<sup>a</sup></b>	Infrared thermographic examination of materials and objects that requires the addition of external thermal stimulation.
<b>Ampligram (image of magnitude)</b>	Image in which each pixel represents the magnitude of Fourier (wavelet) transform at a particular frequency.
<b>Background reflected temperature</b>	The <i>Radiation temperature</i> that is conditioned by the background (ambient) radiation reflected from the surface of a test object.
<b>Blackbody</b>	An ideal body that emits maximal thermal radiation at given temperature and wavelength. Blackbody absorptivity is equal to unity, and reflectivity is zero.
<b>Correlogram</b>	Image that contains pixel-based values of correlation coefficient (correlation is determined between defect and defect-free areas).
<b>Defect characterization (identification, parameters retrieving)</b>	Determining defect parameters by applying inverse algorithms to experimental data.
<b>Depth profiling</b>	Reconstructing in-depth sample structure.
<b>Difference temperature signal</b>	$\Delta T(x, y, \tau) = T(x, y, \tau) - T_{ref}(x, y, \tau)$ , where $T(x, y, \tau)$ is the sample temperature, $T_{ref}(x, y, \tau)$ is the reference temperature.
<b>Differentiated contrast</b>	$Con_{dif}(x,y,\tau^{(n)}) = T(x,y,\tau) - \sqrt{\tau^{(n)}/\tau} T(x,y,\tau^{(n)})$ , where $\tau^{(n)}$ is a normalization time. This term is based on the classic solution for the adiabatic heating of a semi-infinite body.
<b>Dynamic thermal tomography</b>	Tomographic representation created by an analysis of the time versus temperature evolution in a one-sided (front-surface) test.
<b>Early detection time</b>	A time $\tau^*$ when a temperature signal $\Delta T(\tau)$ first exceeds the noise level. Typically, $\tau^*$ is much shorter than the optimum observation time $\tau_m$ .
<b>Flash technique (Pulsed ITT)</b>	Inspection with a thermal pulse whose duration is significantly shorter than the observation time; typically, samples are heated with a pulsed laser or flash tubes (see ▶ Chap 38, “Active Thermography”).
<b>“Flying spot” technique</b>	ITT by heating an object with a localized moving heat source while monitoring the surface temperature at a fixed distance (and time) of the heated point (high scanning speed).
<b>Frequency-domain analysis</b>	Analysis of a temperature frequency spectrum.
<b>Frequency-modulated thermal wave imaging</b>	Thermal wave frequency is modulated in a single test to provide better conditions for the detection of defects at different depths.
<b>Induction (eddy-current) thermography</b>	Thermography that stimulates objects by electromagnetic radiation (radio waves).
<b>Infrared thermographic testing (thermographic testing)<sup>a</sup></b>	Inspection of materials and products using <i>Infrared thermography</i> .
<b>Infrared thermography (IR thermography)<sup>a</sup></b>	Imaging an object by sensing the infrared (thermal) radiation emitted by it.
<b>Infrared (IR) thermogram<sup>a</sup></b>	A thermal map or image of a target where the grey tones or color hues represent the distribution of infrared thermal radiant energy emitted by the surface of the target.
<b>Lock-in IR thermography</b>	Phase-sensitive IR thermography that uses low-frequency thermal stimulation of tested objects.
<b>Local defect resonance (LDR) technique</b>	<i>Ultrasonic IR thermographic</i> inspection by using low-power ultrasonic stimulation with varying frequencies to achieve resonance within a particular defect thus enhancing temperature signals.

(continued)

**Table 2** (continued)

**Maxigram** Image that contains pixel-based maximum values of a selected parameter (*Difference temperature signal*) independent of the time of their appearance.

**Normalized temperature contrast**

$$Con_{norm}(x,y,\tau) = -\frac{T(x,y,\tau)-T_{in}(x,y)}{T^{(n)}(x,y,\tau^{(n)})-T_{in}(x,y)} - \frac{T_{ref}(x,y,\tau)-T_{in}(x,y)}{T_{ref}^{(n)}(x,y,\tau^{(n)})-T_{in}(x,y)}$$

where  $T^{(n)}(x, y, \tau^{(n)})$  is the temperature of a current point at time  $\tau^{(n)}$ , which is assumed to be normalizing. Frequently,  $\tau^{(n)}$  corresponds to the end of heating  $\tau_h$ .  $T_{in}(x, y)$  is the sample initial (ambient) temperature.

**Optimum observation time** A time  $\tau_m$  when a maximum value of a chosen critical parameter appears.

**Passive thermography (passive ITT)<sup>a</sup>** Thermographic technique for inspecting objects or structures by monitoring their emitted thermal radiation, without using an additional energy source for thermal stimulation.

**Phasegram** Image that contains pixel-based values of Fourier (wavelet) transform phase at a particular frequency.

**Photoacoustic method** An ultrasonic NDE method in which acoustic signals are produced by optical illumination of samples.

**Photothermal (optothermal) method** The thermal stimulation of a sample is provided by optical radiation (typically, in a small area thus requiring scanning the heat over the sample).

**Polynomial fitting** Fitting “noisy” temporal evolution functions with polynomial functions (a source sequence of arbitrary length can be replaced with a few images made up of polynomial coefficients).

**Principal component analysis (PCA)** A type of *Singular value decomposition*.

**Pulse phase thermography (PPT)<sup>a</sup>** Processing technique used in *Pulsed thermography* in which data are analyzed in the frequency domain rather than in the time domain.

**Pulsed thermography<sup>a</sup>** Active infrared thermographic inspection technique in which a test sample is stimulated with a pulse of energy and the recorded infrared image sequences are analyzed to enhance defect “visibility” and to characterize defect parameters.

**Radiation temperature** Apparent temperature which corresponds to the power of a recorded thermal radiation. It is equal to the true temperature in the case of a *Blackbody*.

**Reflection mode (front-surface, one-sided test)** Thermal stimulation and temperature monitoring occur on the same surface of a sample.

**Running temperature contrast**  $Con_{run}(x,y,\tau) = \frac{\Delta T(x,y,\tau)}{T(x,y,\tau)-T_{in}(x,y)}$ .

**Singular value decomposition (SVD)** Calculation of a covariation matrix and the eigenvectors of its input data. Eigenvectors are arranged in the order of their magnitude, thus providing the components of an analyzed statistical set. By neglecting low-order components, it is possible to improve results, by reducing the effect of such things as uneven heating.

**Space-domain analysis** Analysis of temperature amplitude parameters.

**Stimulated (forced, active, transient) ITT** Additional thermal stimulation is applied.

**Synthetic signal processing** A patented data processing technique that involves: (1) polynomial fitting; (2) restoring a source function by polynomial coefficients; (3) analysis of temperature derivatives in time.

**Thermal contrast<sup>a</sup>** The degree of detectable temperature difference between adjacent areas that have unequal temperatures at a particular time. Note: Thermal contrast is a processing technique used to enhance defect visibility. In its simplest form, the thermal contrast is computed by calculating the difference between the temperature of the target area and the temperature of a (sound) reference area.

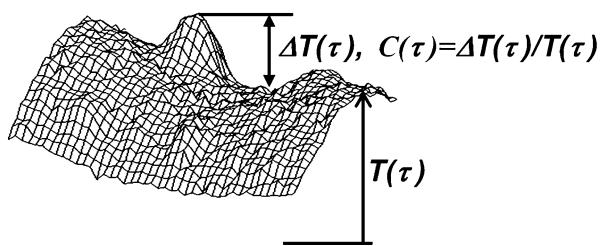
(continued)

**Table 2** (continued)

<b>Thermal tomography</b> <sup>a</sup>	A data processing technique used in pulsed thermography in which data is analyzed in comparison with a particular time of interest, such as the time of maximum thermal contrast.
<b>Thermoelastic stress image</b>	IR thermogram image that shows the thermoelastic stresses.
<b>Thermogram</b>	Image that contains pixel-based values of radiation temperature.
<b>Thermographic signal reconstruction (TSR)</b> <sup>a</sup>	A signal-processing technique for reconstructing and improving time-resolved thermal images, used in pulsed thermography testing and based on polynomial fitting of temperature decay. This is a type of <i>Synthetic signal processing</i> .
<b>Thermography</b>	Analysis of spatial and temporal distribution of thermal parameters (temperature) in objects, typically accomplished by creating corresponding infrared images ( <i>Thermograms</i> ).
<b>Time-domain analysis</b>	Analysis of temperature temporal parameters.
<b>Timegram</b>	Image that contains pixel-based values of a characteristic heat transit time, e.g., <i>Optimum observation time</i> .
<b>Time-resolved infrared radiometry (TRIR)</b>	Analyzing temporal evolution of temperature signals.
<b>Transmission mode (rear surface, two-sided) test</b>	The thermal stimulation and temperature monitoring occur on opposite sides of a sample.
<b>Ultrasonic IR thermography</b>	Thermography in which objects are stimulating by periodic ultrasonic waves.
<b>Ultrasonic burst excitation, sonic IR imaging, thermosonic method</b>	Thermography in which objects are stimulated by pulsed ultrasonic waves.
<b>Vibrothermography</b> <sup>a</sup>	Thermographic technique in which temperature differences are created by mechanical vibrations.

<sup>a</sup>These terms are given in accordance with ISO # 10878 international standard (1st edition, 11 January 2013).

**Fig. 1** Difference temperature signal  $\Delta T$  and running temperature contrast  $C$



stimulation. Additionally, defects can be active or passive. Active defects generate or absorb thermal energy and therefore they can be detected in a passive regime. Passive test samples are at the same temperature as the environment prior to inspection. Therefore, such objects must be heated or cooled to produce defect indications.

In an active ITT mode, the temperature difference between a defect and the surrounding defect-free material changes as a function of the time after heating. Models of active ITT can be classified by: (1) the type of thermal stimulation, (2) the arrangement of the sample and the thermal stimulation source, and (3) the size and shape of stimulated areas.

Thermal stimulation can include both heating and cooling. The highest power density of optical radiation is provided by lasers and Xenon flash tubes (Fig. 2a), while common electrical quartz lamps allow lower-power stimulation. However, the heating rate provided by lasers and flash tubes can be high enough to damage the material being inspected. Quartz lamps can provide a large amount of energy over a relatively large area at a lower rate (seconds) that is less likely to damage materials. And while some test cases require a very fast (short) heating pulse, in others, slower heating provides satisfactory results.

Electrically conductive materials can be heated by induction producing low radiation noise. And by this method, metals may be heated through nonmetallic layers, as shown in Fig. 2b. A particular example of this method is induction (eddy current) IR thermography (Fig. 2c), which involves the use of high frequency currents ( $\sim 100$  kHz).

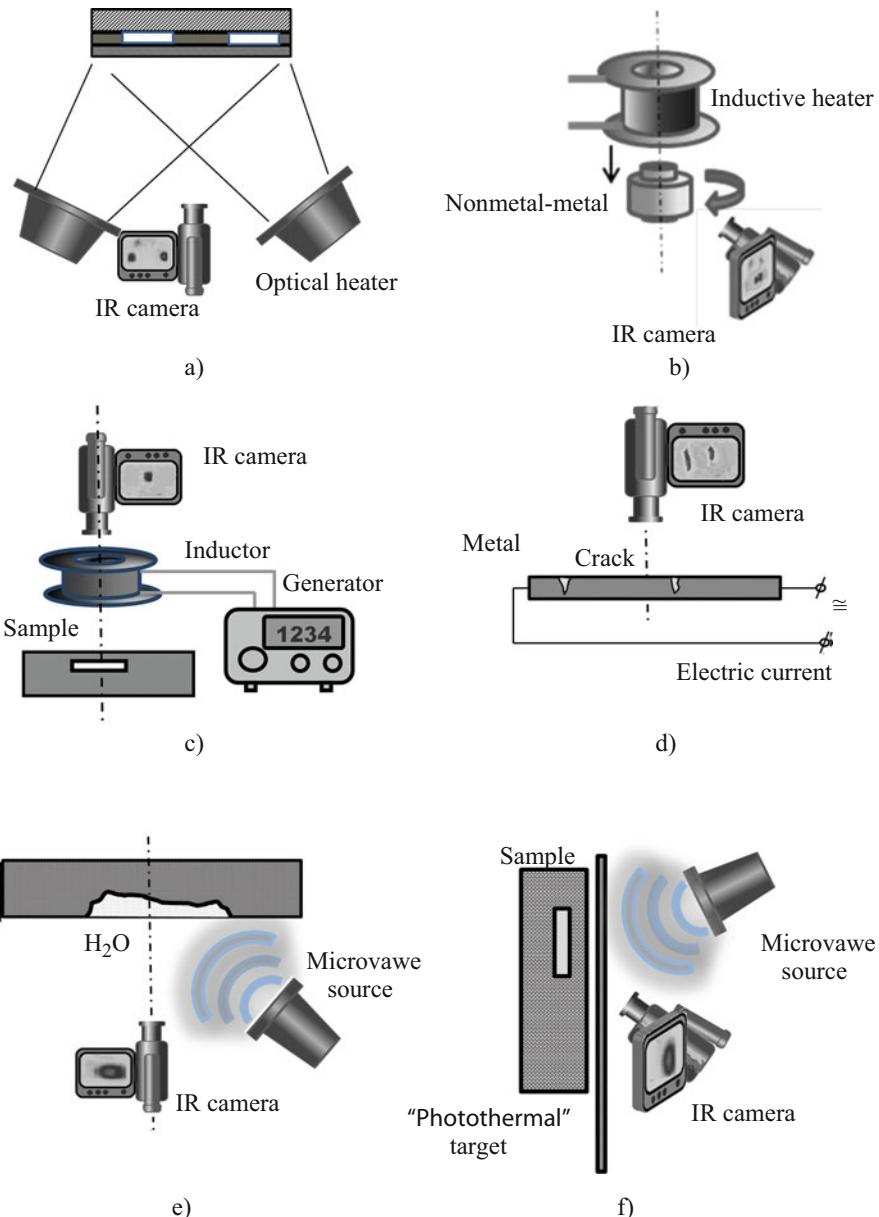
In some cases, metallic samples can be resistively heated by electrical current (Fig. 2d). This technique produces low noise and can be used to detect cracks that are perpendicular to the direction of the current. Note that electrical resistive heating is the basis of the IR testing of the electrical connections in electrical power distribution systems.

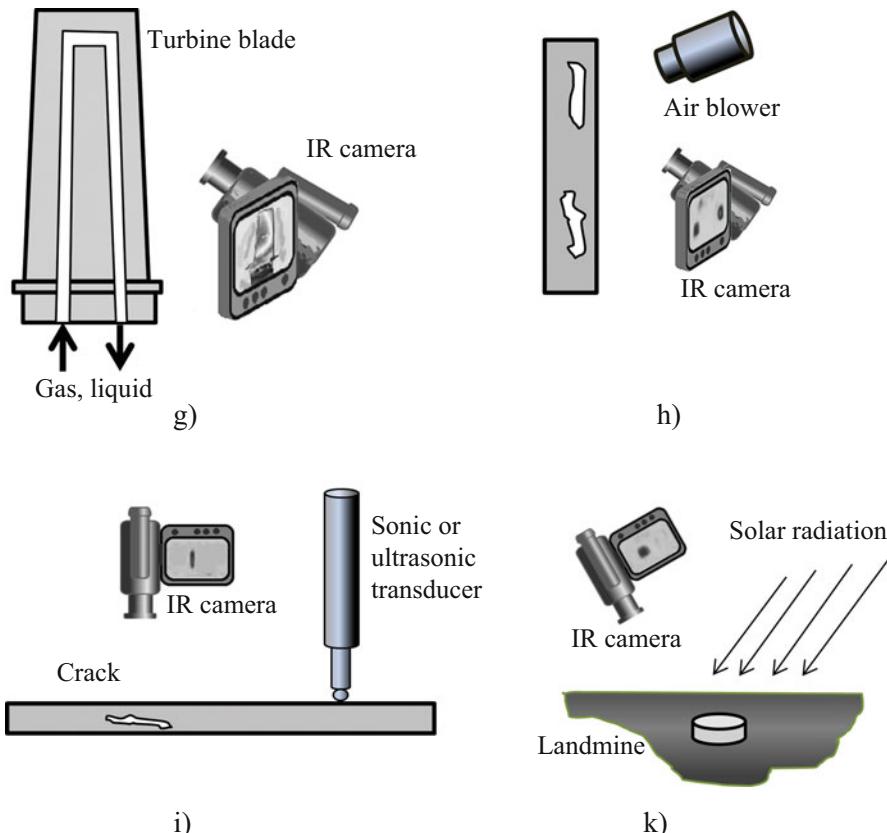
IR thermographic monitoring of materials subjected to microwave (ultra-high frequency) heating can be used to detect moist areas in porous nonmetals (Fig. 2e). Another testing approach combining a microwave source and an IR camera is shown in Fig. 2f. The electromagnetic IR thermography technique involves a photothermal target that is placed close to a test object. An IR imager monitors the target to observe the temperature rise that appears due to the superposition of incident and reflected microwave radiation. This technique has been used for visualizing terahertz radiation.

Convective heating can be implemented by spraying a gas or liquid on or near the surface of a sample (Fig. 2g, h). For example, hot or cold water can be passed through the cooling channels of a turbine blade to thermally stimulate it by convection (Fig. 2g). One-sided heating (Fig. 2h) may be provided by hot air from blowers. In some test cases, samples can be cooled by spraying them with liquid nitrogen. Convective air heating may be as good as or better than optical stimulation because of lower noise, a higher heating efficiency, and the absence of reflected radiation.

The main advantage of mechanical stimulation, or vibrothermography, other than low noise, is the fact that temperature anomalies appear only in defect areas due to friction between defect (crack) surfaces or other phenomena, such as plastic deformation and mechanical hysteresis (Fig. 2i). This heating approach has been very efficient in detecting defects in fiber-reinforced composites excited by ultrasonic transducers.

In other passive heating techniques, the use of specialized heating devices may be unnecessary. Objects such as antipersonnel mines, bridge decks, decorative coatings on building envelopes, etc. are subjected to natural heating by solar radiation. The power of such radiation on a cloudless day may reach about  $1 \text{ kW/m}^2$  in the middle latitudes, varying periodically during a day-night (diurnal) cycle (Fig. 2k).

**Fig. 2** (continued)



**Fig. 2** ITT procedures: (a) optical heating, (b) eddy current heating (low frequency), (c) eddy current heating (high frequency), (d) electric current heating, (e) microwave heating, (f) electromagnetic IR thermography, (g) internal heating with gas/liquid, (h) external blowing with air, (i) mechanical heating, (k) natural heating

## Basics of Heat Transfer

### Heat Conduction, Convection, and Radiation

If temperature gradients exist in media, thermal energy can be transported by three mechanisms: heat conduction (mostly in solids), convection (in liquids and gases), and radiation (mostly through gases and vacuum). The thermal flux density  $Q$  between two points with different temperatures  $T_1$  and  $T_2$  is governed by Fourier's law:

$$Q = -\lambda (\partial T / \partial x). \quad (1)$$

The thermal conduction coefficient  $\lambda$  in W/(m·K) determines the density of the heat flux in a solid if the temperature gradient is unity.

Convection involves the mass movement of gas or liquid molecules over large distances. The surface is cooled down (or warmed up) by convection according to Newton's law:

$$Q = \alpha_{cv}(T_s - T_{amb}). \quad (2)$$

where  $\alpha_{cv}$  is the coefficient of convective heat exchange,  $T_s$  and  $T_{amb}$  are the temperatures of the body and the convective medium (fluid), respectively.

Radiation is a flux of electromagnetic quanta emitted by all physical bodies having a temperature above absolute zero (0 K). Radiant heat exchange between two bodies is described by the Stephan-Boltzmann's law:

$$Q_{rd} = \sigma\Gamma\varepsilon_{red}(T_1^4 - T_2^4), \quad (3)$$

where  $Q_{rd}$  is the radiant heat flux,  $\sigma$  is the Stephan-Boltzmann constant ( $\sigma = 5.67 \cdot 10^{-8} \text{W}/(\text{m}^2 \cdot \text{K}^4)$ ),  $\Gamma$  is a dimensionless geometric factor, and  $\varepsilon_{red}$  is the reduced emissivity. If radiation heat exchange occurs between two plates ( $T_1 > T_2$ ):

$$Q_{rd} \approx \varepsilon_{red}\sigma(T_1^4 - T_2^4), \quad (4)$$

where  $\varepsilon_{red} = 1/\left(\frac{1}{\varepsilon_1} + \frac{1}{\varepsilon_2} - 1\right)$ ,  $\varepsilon_1$  and  $\varepsilon_2$  are the emissivities of the plates. Emissivity is an optical characteristic of the material which represents the ratio between the thermal radiation of a particular material and the thermal radiation of a black body having the same temperature.

The heat exchange between a warm solid body ( $T_s$ ) and a gas or liquid surrounding ( $T_{amb}$ ) occurs due to both convection and radiation:

$$Q = Q_{cv} + Q_{rd} = \alpha_{cv}(T_s - T_{amb}) + \sigma\varepsilon_{red}(T_s^4 - T_{amb}^4). \quad (5)$$

In ITT, a temperature difference  $T_s - T_{amb}$  is typically small and emissivities are typically high ( $\varepsilon_{red} \approx 1$ ). Therefore:

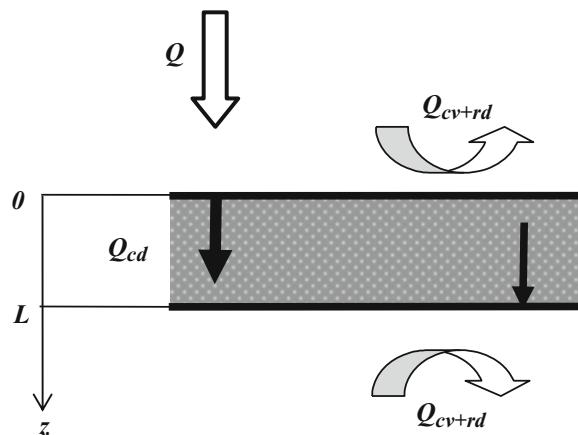
$$\sigma\varepsilon_{red}(T_s^4 - T_{amb}^4) \approx 4\sigma\left(\frac{T_s + T_{amb}}{2}\right)^3(T_s - T_{amb}) = \alpha_{rd}(T_s - T_{amb}), \quad (6)$$

where  $\alpha_{rd}$  is the heat exchange coefficient conditioned by radiation.

Expressions (2) and (6) become identical with the introduction of the combined heat exchange coefficient  $\alpha = \alpha_{cv} + \alpha_{rd}$ . The value of  $\alpha$  depends on the object shape and its orientation in space, as well as on the temperature difference  $T_s - T_a$ .

In calculus, it is often assumed that  $\alpha \approx 10 \text{ W}/(\text{m}^2 \cdot \text{K})$  at the ambient temperature. Absence of surface heat exchange ( $\alpha_{cv} = \alpha_{rd} = 0$ ) specifies adiabatic boundary

**Fig. 3** Boundary conditions on the test object surface



conditions. In ITT, such conditions appear in the inspection of metals and thin nonmetals, in particular, at short observation times.

## Boundary Conditions

On a heated surface, the boundary condition is constituted by the sample external heating and its cooling with convection and radiation (Fig. 3). The corresponding expression represents the balance between three heat fluxes: the heating flux  $Q$ , the convective/radiant flux  $Q_{cv+rd}$ , and the conductive flux  $Q_{cd}$  which penetrates the sample:

$$Q = Q_{cv+rd} + Q_{cd}. \quad (7)$$

A typical presentation of Eq. (7) is:

$$-\lambda \frac{\partial T(z=0)}{\partial z} = Q - \alpha [T(z=0) - T_{amb}], \quad (8)$$

where  $z = 0$  specifies the front surface. On the rear surface ( $z = L$ ,  $L$  is the sample thickness), there is no heating flux, therefore, there is the balance of only two heat fluxes:

$$-\lambda \frac{\partial T(z=L)}{\partial z} = \alpha [T(z=L) - T_{amb}], \quad (9)$$

where the term  $Q_{cv+rd}$  changes its sign due to the reversed direction of the heat flux with respect to the front surface (Fig. 3).

In active ITT, the incident heating flux is typically much more powerful than heat fluxes caused by convection and radiation. Therefore, in this case, the heat exchange turns to be adiabatic, and the corresponding heat transfer solutions become simple.

## Heat Transfer in Defects

In ITT, hidden defects are often considered as thin air gaps located perpendicularly to the main heat flux. Heat transfer in such defects occurs through conduction, convection, and radiation. Each of the three heat flux components is determined by the temperature gradient  $\Delta T = T_1 - T_2$  between defect edges:

$$\begin{aligned} Q_{cd} &= \lambda_d (T_2 - T_1)/d \text{ (conduction),} \\ Q_{cv} &= \lambda_d^c (T_2 - T_1)/d \text{ (convection),} \\ Q_{rd} &\approx \varepsilon_{red}\sigma (T_1^4 - T_2^4) \approx 4\varepsilon_{red}\sigma T_{0.5}^3 (T_1 - T_2) \text{ (radiation).} \end{aligned} \quad (10)$$

Here  $d$  is the defect thickness,  $T_{0.5} = \frac{T_1+T_2}{2}$ , K, and the convective component of thermal conductivity  $\lambda_d^c$  is determined by the product of two criteria known in the heat transfer theory as Grashof and Prandtl (Gr Pr):

$$\lambda_d^c = (\zeta - 1) \lambda_d. \quad (11)$$

If  $10^3 < \text{Gr } \text{Pr} < 10^6$ ,

$$\zeta = 0.105 (\text{Gr Pr})^{0.3} \quad (12)$$

where  $\text{Gr} = \frac{gd^3}{\nu^2} \beta (T_1 - T_2)$ ,  $\text{Pr} = \nu/a$ ,  $g$  is the acceleration of free gravity,  $\text{m/s}^2$ ,  $\nu = \mu/\rho$  is the coefficient of air kinematic viscosity,  $\text{m}^2/\text{s}$ ,  $\mu$  is the coefficient of air dynamic viscosity,  $\text{N}\cdot\text{s}/\text{m}^2$ ,  $a = \frac{\lambda}{C\rho}$  is thermal diffusivity,  $\text{m}^2/\text{s}$ ,  $C$  is the specific heat,  $\text{J}/(\text{kg}\cdot\text{K})$ ,  $\rho$  is the density,  $\text{kg}/\text{m}^3$ ,  $\beta = \frac{1}{t_{0.5} + 273}$ , and  $t_{0.5} = (t_1 + t_2)/2$ ,  $^\circ\text{C}$ .

In thin air-filled defects, the conductive heat flux essentially dominates over the heat fluxes conditioned by convection and radiation. For example, if the defect thickness is  $d < 250 \mu\text{m}$ , the increase in the value of air effective thermal conductivity is under 1.5% to compare to the handbook value for free air.

The situation becomes different if the air gap thickness exceeds 5–20 mm, as in honeycomb cells. In such gaps, a significant convection may appear. For example, in honeycomb cells, the effective thermal conductivity of air increases from 0.07 to 0.28  $\text{BT}/(\text{m}\cdot\text{K})$ .

By assuming pure heat conduction through defects, there are two types of boundary conditions describing the contact between layers.

In general, on the boundary between layers  $i$  and  $i + 1$ , one should assume a condition of the continuity of both heat fluxes and temperatures:

$$\lambda_i \frac{\partial T_i}{\partial z} = \lambda_{i+1} \frac{\partial T_{i+1}}{\partial z}, \quad T_i = T_{i+1}. \quad (13)$$

While assuming (13), a differential heat conduction equation must be solved in both the host material and in each defect by matching the corresponding solutions at the boundaries. In this case, the temperature distribution in a sample will depend on layer thermal conduction and thermal diffusivity (heat capacity), therefore, the corresponding defects are called *thermally capacitive*.

If the heat capacity of the  $i + 1$  layer can be neglected, its main thermal parameter becomes the thermal resistance  $R_{i+1} = d_{i+1}/\lambda_{i+1}$ ; and the corresponding defects are called thermally resistive. At the boundary between the host material and a resistive defect, the temperature changes abruptly and the heat flux remains constant:

$$-\lambda_i \frac{\partial T_i}{\partial z} = -\lambda_{i+1} \frac{\partial T_{i+1}}{\partial z}; \quad T_{i+1} - T_i = R_{i+1} \frac{\partial T_{i+1}}{\partial z}. \quad (14)$$

## Differential Equation of Heat Conduction

Nonstationary (transient) temperature distributions in solids with internal heat sources are governed by the parabolic differential equation of heat conduction:

$$\frac{\partial}{\partial x} \left( \lambda_x \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left( \lambda_y \frac{\partial T}{\partial y} \right) + \frac{\partial}{\partial z} \left( \lambda_z \frac{\partial T}{\partial z} \right) + w(x,y,z) = C\rho \frac{\partial T}{\partial \tau}, \quad (15)$$

where  $w(x, y, z)$  is the rate of energy generation per unit volume in the medium ( $\text{W/m}^3$ ),  $C$  and  $\rho$  are specific heat capacity ( $\text{J/(kg}\cdot\text{K)}$ ) and material density ( $\text{kg/m}^3$ ), respectively. Eq. (15) reflects the principle of energy conservation in media where heat energy is generated and transported by diffusion.

The anisotropic nature of heat diffusion in Eq. (15) is expressed by thermal conduction coefficients  $\lambda_x, \lambda_y, \lambda_z$ . In the case of an isotropic material:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} + \frac{w}{\lambda} = \frac{1}{a} \frac{\partial T}{\partial \tau}, \quad (16)$$

where  $a = \lambda/(C\rho)$  is the thermal diffusivity ( $\text{m}^2/\text{s}$ ).

When there is no internal heat source ( $w=0$ ), the following known equation appears:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} = \frac{1}{a} \frac{\partial T}{\partial \tau}, \text{ or } \nabla^2 T = \frac{1}{a} \frac{\partial T}{\partial \tau}. \quad (17)$$

## Analytical Solutions to ITT Problems

Classical heat conduction theory supplies many analytical solutions describing heating of bodies with a simple geometry, such as a plate, cylinder, and sphere (Carslow and Jaeger 1959). Finite-size subsurface defects disturb flowing of the regular heat flux in solids thus causing appearance of local temperature anomalies. Therefore, typical ITT problems are essentially 3D and they are being normally solved by using numerical models. In defect-free areas, one may use the corresponding analytical solutions some of which are presented in Table 3.

## Basic Relationships in ITT

Consider a classical scheme of pulsed TNDE depicted in Fig. 4. A test object contains an air-filled defect at the depth  $l$  (Fig. 3a). Defect lateral size is  $D_x \times D_y$ , and defect thickness along the  $z$ -coordinate is  $d$ . The object is stimulated with a heat pulse on the front ( $F$ ) surface. In a one-sided test procedure, the temperature is monitored on the  $F$ -surface, while a two-sided test involves temperature recording on the rear ( $R$ ) sample surface. The result of any test is a sequence of IR images where pixel-based temperature evolutions are analyzed to ensure defect detection. These evolutions are shown in Fig. 4b, c for the two sample surfaces along with the difference temperature signals  $\Delta T$ . The most important test parameters are maximal signals  $\Delta T_m$  and their corresponding (optimum) observation times  $\tau_m$ . Since the accepted definition of the difference temperature signal is  $\Delta T = T_d - T_{nd}$ ,  $\Delta T$  signals are positive on the  $F$ -surface and negative on the  $R$ -surface. In other words, defects with lower thermal conductivity appear warmer than the background in a one-sided test and colder in a two-sided test.

The dependencies of the above-mentioned parameters on defect depth and size are presented in Fig. 5. On the front surface, with greater  $l$ , the temperature signal  $\Delta T_m$  declines and optimum observation time occurs later, while in two-sided TNDE both  $\Delta T_m$  and  $\tau_m$  values are weakly dependent on defect depth (Fig. 5a, b). In the latter case, the decisive parameter is sample thickness  $L$  and thermal properties.

The influence of defect thickness  $d$ , or defect thermal resistance  $R_d = d/K_d$  ( $K_d$  is the defect thermal conductivity), on temperature signals and observation times reveals saturation in a  $F$ -surface test. In a two-sided procedure, thicker defects are characterized by higher  $\Delta T_m$  and longer  $\tau_m$  values (Fig. 5c, d).

Influence of defect lateral size  $D$  on significant test parameters is conditioned by lateral heat diffusion. The corresponding dependence is also characterized by saturation in the case of extended defects (Fig. 5e, f). Approximately, one can assume that the lateral heat flux is negligible in nonmetals (composites) if  $D/l > 5$  and in metals if  $D/l > 10$ . In the literature, the following rule-of-thumb can be encountered:  $D/l > 2$ ; in fact, this condition specifies only a sharp drop of the  $\Delta T_m$  versus  $D/l$  relationship.

Realistic estimates of  $\Delta T_m$  and  $\tau_m$  can be obtained by numerically solving 3D TNDE problems. For very approximate calculations, which provide an order of

**Table 3** Selected analytical solutions in ITT. (Adapted from (Vavilov 2017))

Problem	Analytical solution
Semi-infinite body is heated with uniformly distributed heat flux described by step-function	$\frac{T}{Q/\lambda} = 2 \left( \frac{a\tau}{\pi} \right)^{1/2}$ (also applicable to a plate at shorter times)
Point-like heating of semi-infinite body	$\frac{T}{W} = \frac{1}{4(\pi a\tau)^{3/2}} e^{-\frac{\tau^2}{4a\tau}}; W' = W/\rho C$ (also applicable to a plate at shorter times)
Semi-infinite body is heated with uniformly distributed heat flux described by Dirac pulse (flash)	$\frac{T}{W} = \frac{1}{e\sqrt{\pi\tau}}; e = \sqrt{C\rho\lambda}$
Semi-infinite body is heated with uniformly distributed heat flux described by harmonic function	$\frac{T(\tau)}{Q_m} = \frac{1}{\lambda\sqrt{\omega/a}} e^{-z\sqrt{\omega/2a}} \cos \left[ \omega\tau - \left( z\sqrt{\omega/2a} \right) - \Theta \right];$ $\Theta = 45^\circ$ $z$ is in-depth coordinate; $z = 0$ corresponds to the surface.
Plate is heated with uniformly distributed heat flux described by Dirac pulse (flash), front surface solution	$\frac{T}{W'} = 1 + 2 \sum_{n=1}^{\infty} e^{-n^2\pi^2 Fo}, W' = \frac{Wa}{\lambda L}$
Plate is heated with uniformly distributed heat flux described by Dirac pulse (flash), rear surface solution	$\frac{T}{W'} = 1 + 2 \sum_{n=1}^{\infty} (-1)^n e^{-n^2\pi^2 Fo}$
Plate is heated with a moving linear heat source $L$ – plate thickness, $V$ – heat source velocity, $q$ – heat source power in $\text{W m}^{-1}$ , $x$ – coordinate in movement direction (coordinate origin corresponds to position of linear heat source)	$T(x,z) = \frac{q}{LVC\rho} \left[ e^{-V\frac{ x +x}{2a}} + 2V \sum_{n=1}^{\infty} \cos \left( \frac{n\pi z}{L} \right) \frac{e^{\frac{-Vx+ x \sqrt{(2n\pi a/L)^2+V^2}}{2a}}}{\sqrt{(2n\pi a/L)^2+V^2}} \right]$ If $V \gg 2a/L$ : $T \approx \frac{q}{LVC\rho}$
Coating on semi-infinite substrate is heated with uniformly distributed Dirac pulse (flash)	$T = \frac{W}{e_c\sqrt{\pi\tau}} \left[ 1 + 2 \sum_{n=1}^{\infty} (-G)^n \exp \left( -n^2 \frac{L_c^2}{a_c\tau} \right) \right]$

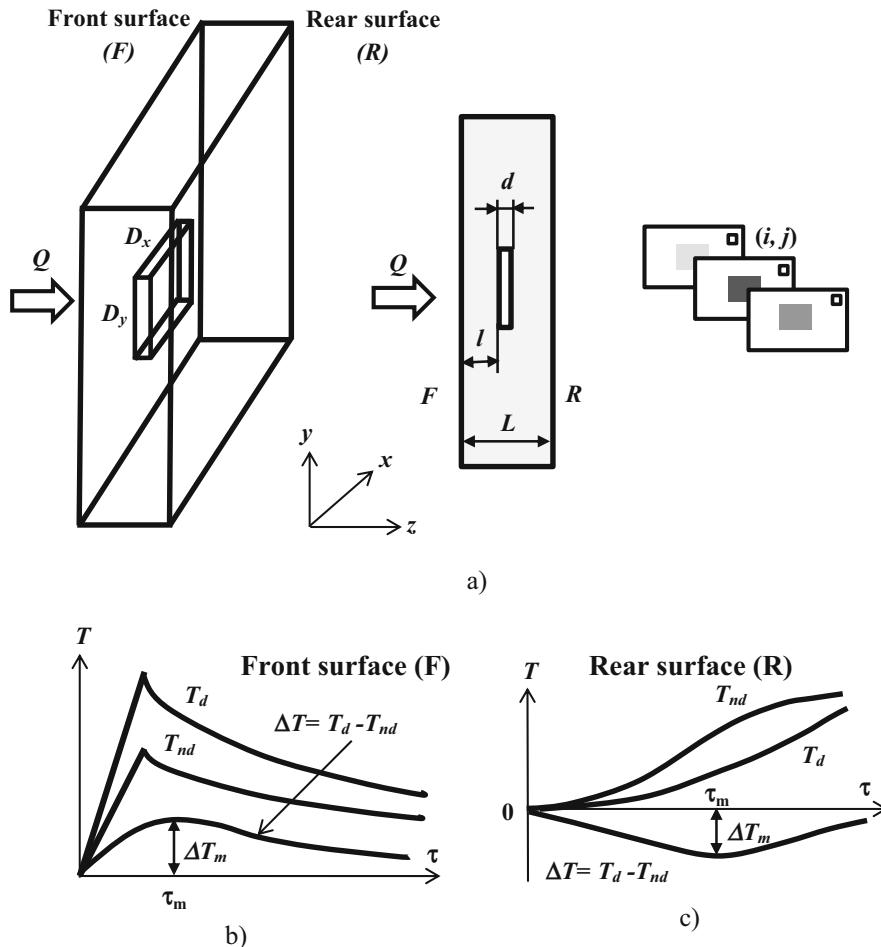
Notation:  $T$  – sample excess temperature,  $Q$  – absorbed heat flux power,  $\lambda$  – thermal conductivity,  $a$  – thermal diffusivity,  $\tau$  – time;  $W$  – absorbed heat flux energy,  $\rho$  – density,  $C$  – heat capacity,  $e$  – thermal effusivity,  $z$  – in-depth coordinate,  $L$  – plate thickness,  $Fo$  – Fourier number (dimensionless time),  $G = (e_c - e_s)/(e_c + e_s)$ , “ $c$ ” specifies coating. “ $s$ ” – substrate.

magnitude of TNDE significant parameters, the following formulas are recommended for nonconductive defects.

In a front-surface test procedure:

$$\Delta T(\tau) = T(l, \tau) - T(L, \tau), \tau_m = \frac{l^2}{a}; \quad (18)$$

in a two-sided test procedure:



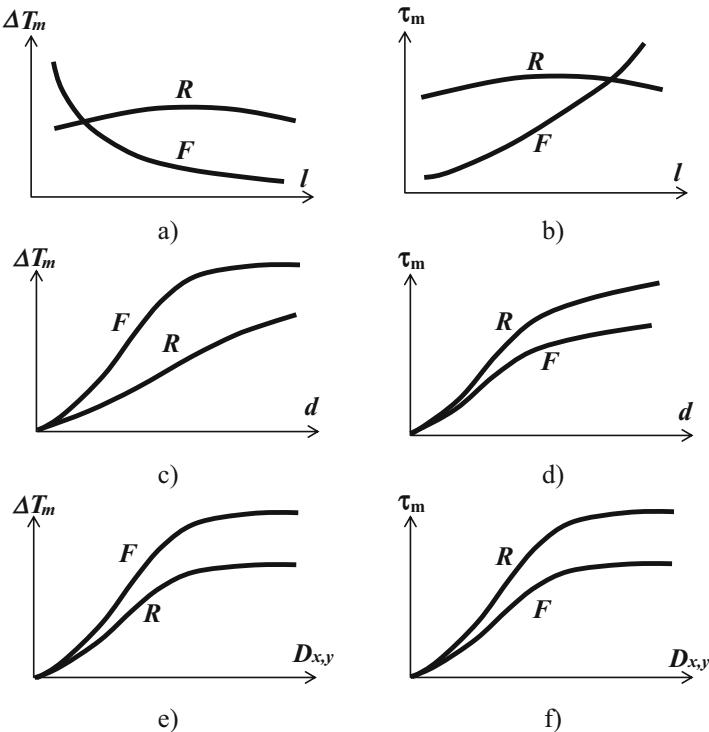
**Fig. 4** TNDE scheme. (a) Test sample with defect ( $l$ -defect depth,  $d$ -defect thickness,  $D$ -defect lateral size). (b) Temperature vs. time (front surface). (c) Temperature versus time (rear surface)

$$\Delta T(\tau) = T_{nd}(\tau), \quad \tau_m = 0.14L^2/a. \quad (19)$$

Here  $T(l, \tau)$  and  $T_{nd}(\tau)$  can be calculated by using the corresponding solutions from Table 3. The estimates above become more accurate for defects with very low thermal conductivity.

## Comparing Test Procedures and Processing Algorithms

Signal-to-noise ratio ( $SNR$ ) is a universal criterion used for comparing NDE procedures and data processing algorithms. As shown in Fig. 6, left, one should choose two areas on the sample surface: non-defect (ND) and defect (D) and



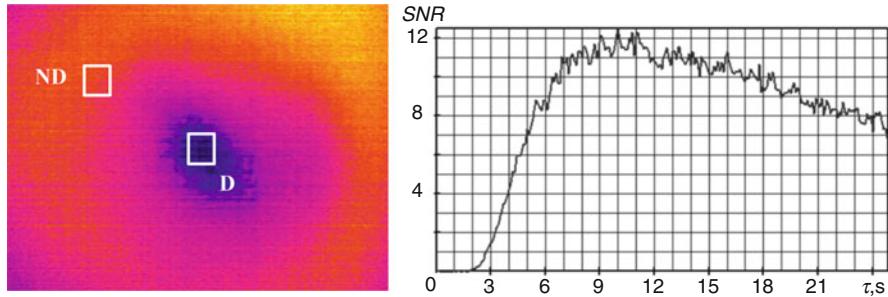
**Fig. 5** Basic relationships in TNDE ( $\Delta T_m$ -maximal temperature signal,  $\tau_m$ -optimum observation time,  $F$ -front surface,  $R$ -rear surface,  $l$ -defect depth,  $d$ -defect thickness,  $D$ -defect lateral size)

determine their simple statistical parameters: mean signals  $\overline{T}_{nd}$ ,  $\overline{T}_d$  and the standard deviation in a non-defect area  $\sigma_{nd}$ . Then, the corresponding  $SNR$  value can be calculated by:

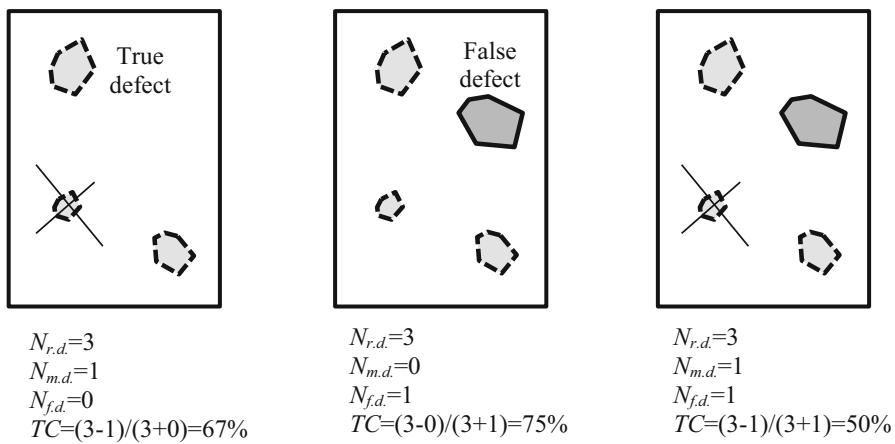
$$SNR = \frac{|\overline{T}_d - \overline{T}_{nd}|}{\sigma_{nd}}. \quad (20)$$

Unfortunately, this criterion is dependent on size and location of both defect and non-defect areas. Since heating is often uneven, it is recommended to place a non-defect area close to a visible defect border that makes test results even more dependent on the thermographer. It is worth mentioning that  $SNR$  is a function of time, therefore, the optimum observation time is characterized by a maximum  $SNR$  value (see the plot in Fig. 6, right).

Another, not very common comparison tool is the Tanimoto criterion that is typically applied to reference samples with plentiful defects evaluated by different operators (Grinzato et al. 1995):



**Fig. 6** Determining signal-to-noise ( $SNR$ ) ratio (impact damage in 4.8 mm-thick CFRP sample, flash heating, two-sided ITT)

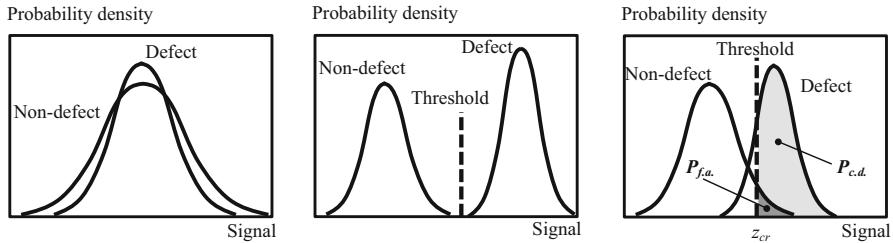


**Fig. 7** Determining Tanimoto criterion by counting defect indications

$$TC = \frac{N_{r.d.} - N_{m.d.}}{N_{r.d.} + N_{f.d.}}, \quad (21)$$

where  $N_{r.d.}$  is the number of real defects in a sample,  $N_{m.d.}$  is the number of missed defects, and  $N_{f.d.}$  is the number of false defects. The use of Eq. (21) is illustrated with three inspection cases in Fig. 7 where a thermographer could identify defect indications which may be conditioned by either real defects or noisy (e.g., emissivity) spots. Note that  $TC$  strongly depends on thermographer's experience and becomes equal to 100% only if  $N_{m.d.} = N_{f.d.} = 0$ . All other cases, such as shown in Fig. 7, produce  $TC < 100\%$ .

A more correct approach is related to the analysis of histograms which represent distributions of pixel amplitudes in both defect and defect-free areas. Three typical cases are shown in Fig. 8. If the two distributions are



**Fig. 8** Signal histograms in NDE

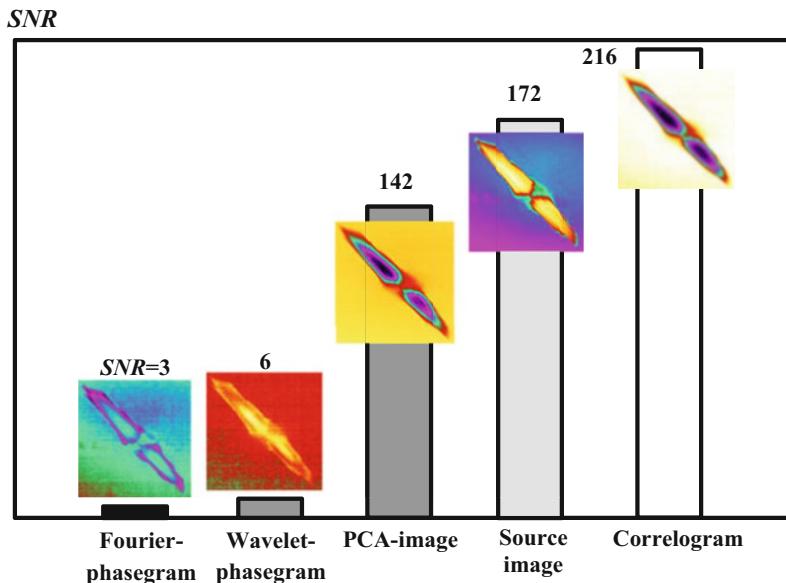
very close, such defects cannot be detected (Fig. 8, left). An opposite case is presented in Fig. 8, center, where “defect” and “non-defect” histograms are well separated thus providing the probability of correct detection  $P_{c.d.} = 100\%$  and the probability of false alarm  $P_{f.a.} = 0$ . In this case, a decision-making threshold can be easily located between the distributions. Finally, a more practical case is shown in Fig. 8, right, where the two distributions partially cross each other. Note that  $P_{c.d.}$  is numerically equal to the area in the “defect” distribution which is located to the right from a chosen threshold  $z_{cr}$  (Fig. 8, right). In turn,  $P_{f.a.}$  is numerically equal to the area in the “non-defect” distribution which is located to the right from a chosen threshold  $z_{cr}$ . According to the Neumann-Pearson strategy of decision-making, a value of  $P_{f.a.}$  is to be fixed, e.g., at the 5% level, thus automatically defining the corresponding value of  $P_{c.d.}$ .

In ITT, the statistics of noise is scarcely studied. If a decision-making parameter (temperature, phase, etc.) is governed by the normal distribution, the theory of statistical decisions provides the following expressions:

$$\begin{aligned} P_{c.d.} &= 1 - \Phi \left( Z = \frac{\bar{U}_d - U_{thr}}{\sigma_{nd}} \right); \\ P_{f.a.} &= \Phi \left( Z = \frac{U_{thr} - \bar{U}_{nd}}{\sigma_{nd}} \right), \end{aligned} \quad (22)$$

where  $\Phi(Z) = \frac{1}{2\sqrt{\pi}} \int_Z^{\infty} e^{-x^2/2} dx$  is the tabulated Laplace integral,  $\bar{U}_d$ ,  $\bar{U}_{nd}$  are the mean values of the signal (significant parameter) in the non-defect and defect areas, respectively, and  $U_{thr}$  is the decision-making threshold. For example, to obtain  $P_{f.a.}=5\%$  one has to set the decision-making threshold equal to  $\bar{U}_{nd} + 1,6 \sigma_{nd}$ .

When comparing NDE methods in practice, one often uses the dependence of  $P_{c.d.}$  on defect size called the probability of detection (POD). This parameter is usually averaged by some operators thus taking into account a human factor. An example of applying such an approach to the inspection of the thermal protection of the US space shuttle X-33 by using ultrasonic, IR thermographic and holographic NDE is reported in Taylor and Dupont (1998).



**Fig. 9** Efficiency of data processing algorithms in TNDE (impact damage in graphite epoxy composite)

## Data Processing Algorithms

Most of the available data processing algorithms are pixel-based, being applied to  $T(i,j,k)$  functions (here  $i,j$  are the pixel coordinates, and  $k$  is the image number in a sequence). In general, such algorithms are supposed to enhance the signal-to-noise ratio or improve defect visibility. However, in some cases, results of image evaluation obtained by thermographers may be different from those prompted by the corresponding SNR values. The example in Fig. 9 illustrates that, in the detection of impact damage in graphite epoxy composites, some known algorithms do not guarantee SNR values higher than those provided by optimal source images. This may take place in the case of complicated defects, such as multiple delaminations/cracking occurring at different depths. However, particular algorithms can be helpful to stress some defect features (compare visually the images in Fig. 9 characterized by different SNR values).

## References

- Almond D, Patel P (1996) Photothermal science and techniques. Chapman & Hall, London
- Balageas DL, Kapez J-C, Cielo P (1986) Pulsed photo-thermal modeling of layered materials. *J Appl Phys* 59(2):348–357
- Beller WS (1965) Navy sees promise in infrared thermography for solid case checking. *Missiles Rockets* 22:1234–1241

- Breitenstein O, Warta W, Langekamp M (2010) Lock-in thermography. In: Springer series in advanced microelectronics, vol 10. Springer, Berlin
- Busse G (1985) Imaging with optically generated thermal waves. In: Physical acoustics, vol 43. Academic, London, pp 403–478
- Carslow HS, Jaeger TS (1959) Conduction of heat in solids. Oxford University Press, Oxford, UK
- Cramer K, Winfree W, Hodges K, Koshti A, Ryan D, Reinhard WW (2006) Status of thermal NDE of space shuttle materials at NASA. In: Proceedings of SPIE "Thermosense-XXVIII", vol 6205, 62051B1-9
- Gorril WS (1949) Industrial high-speed infrared pyrometer to measure the temperature of a soldered seam on a tin can. *Electronics* 22:112–115
- Green DR (1965) Emissivity-independent infrared thermal testing method. *Mater Eval* 23(2):79–85
- Grinzato E, Vavilov V, Bison PG, Marinetti S (1995) Methodology of processing experimental data in transient thermal NDE. In: Proceedings of SPIE "Thermosense-XVII", vol 2473, pp 62–63
- Kush DV, Rapoport DA, Budadin ON (1988) Inverse problem of automated thermal NDE. *Defectoscopiya* (Soviet J. NDE) 5:64–68 (in Russian)
- Leslie JR, Wait JR (1949) Detection of overheated transmission line joints by means of a bolometer. *Trans Am Inst Electr Eng* 68:64–68
- Luong MP (1992) Infrared thermography of fatigue in metals. In: Proceedings of SPIE "Thermosense-XIV", vol 1682, pp 222–232
- Maillet D, Andre S, Batsale J-C et al (2000) Thermal quadrupoles: solving the heat equation through integral transforms. Wiley, Chichester
- Maldague X, Marinetti S (1996) Pulse phase infrared thermography. *J Appl Phys* 79:2694–2698
- Mandelis A (2001) Diffusion-wave fields. Chap. 9. In: Mathematical methods and green functions. Springer, New York
- Nichols JT (1935) Temperature measuring. US Patent 2,008,793
- Parker RC, Marshall PR (1948) The measurement of the temperature of sliding surfaces with particular references in railway brake blocks and shoes. *Proc Inst Mech Eng* 158:209–212
- Pettersson B, Bengt A (1980) Thermography. Testing of the thermal insulation and airtightness of buildings. Swedish Council for Building Research, Stockholm
- Popov Yu A, Karpelson AE, Strokov VA (1976) Thermal NDE of multi-layer structures. *Defectoscopiya* (Soviet J NDE) 3:76–81 (in Russian)
- Rajic N (2002) Principal component thermography for flaw contrast enhancement and flaw depth characterization in composite structures. *Compos Struct* 58(4):521–528
- Shepard S (2001) Advances in pulsed thermography. In: Proceedings of SPIE "Thermosense-XXIII", vol 4360, pp 511–515
- Taylor JO, Dupont HM (1998) Inspection of metallic thermal protection systems for the X-33 launch vehicle using pulsed infrared thermography. In: Proceedings of SPIE "Thermosense-XX", vol 3361, pp 301–310
- Vavilov VP (2007) Pulsed thermal NDE of materials: back to basics. *Nondestruct Test Eval* 22(2–3):177–197
- Vavilov VP (2014) Thermal NDE: historical milestones, state-of-the-art and trends. *QIRT J* 11(1):66–83
- Vavilov VP (2017) Thermal nondestructive testing of materials and products: a review. *Rus J NDT* 53(10):707–730
- Vavilov VP, Burleigh DD (2015) Review of pulsed thermal NDE: physical principles, theory and data processing. *NDE & E Int* 73:28–52
- Vavilov V, Taylor R (1982) Theoretical and practical aspects of the thermal NDE of bonded structures. In: Sharpe R (ed) Research technique in NDE, vol 5. Academic, London
- Vernotte P (1937) Mesure de la conductibilité thermique des isolants. méthode de toushau. *Chaleur Ind* 208:331–337 (in French)
- Vollmer M, Möllmann K-P (2010) Infrared thermal imaging: fundamentals, research and applications. Wiley-VCH, Weinheim



# Passive Thermography, Thermal Imaging

37

Helmut Budzier and Gerald Gerlach

## Contents

Basics .....	1372
Radiation Laws .....	1373
IR Characteristics of Bodies .....	1376
Basic Law of Photometry .....	1381
Sensors .....	1382
Sensor Characteristics .....	1382
Thermal Sensors .....	1383
Photon Sensors .....	1388
Comparison of Thermal and Photon Sensors .....	1393
Thermal Imaging Cameras .....	1394
Passive Thermographic Techniques .....	1397
Preventive Maintenance .....	1397
Testing Electronic Components and Assemblies .....	1398
Process Monitoring .....	1400
References .....	1400

### Abstract

Thermography is an imaging technique designed to use the thermal radiation of an object to determine its characteristics. The surface radiation of an object is measured. With the use of Planck's Radiation Law and knowing the emissivity, it is possible to calculate the surface temperature.

The emissivity of real bodies is explained on the basis of the physical laws governing the emission of radiation. The infrared radiation sensors required for the detection are presented below. These are divided into thermal sensors and photon sensors. Thermal sensors detect the radiation energy, while

---

H. Budzier · G. Gerlach (✉)

Institute of Solid State Electronics, Technische Universität Dresden, Dresden, Germany  
e-mail: [Helmut.Budzier@tu-dresden.de](mailto:Helmut.Budzier@tu-dresden.de); [gerald.gerlach@tu-dresden.de](mailto:gerald.gerlach@tu-dresden.de)

photon sensors record the photon flux. The structure and functionality of thermal imaging cameras are set out. Finally, thermographic techniques for preventive maintenance, testing electronic components, and process monitoring are presented.

---

## Basics

In accordance with Planck's Radiation Law, each object with a temperature above absolute zero emits radiation predominantly in the infrared (IR) spectral range. This radiation is dependent on the temperature and the emission characteristics of the radiating object. By measuring this IR radiation, it is possible to determine both the presence of a radiating object (say, in the dark) and – if the emission characteristics are known – the temperature of the object without contact with it.

In addition to the non-contact temperature imaging (thermography) mentioned above, the main areas of application of infrared metrology are to be found in gas detection as well as in motion, presence, direction, and position detector technology.

The benefits provided by thermography include:

- No energy is extracted from the object measured.
- Both objects in motion and extremely small objects can be measured.
- The measurement can be made from a large distance.

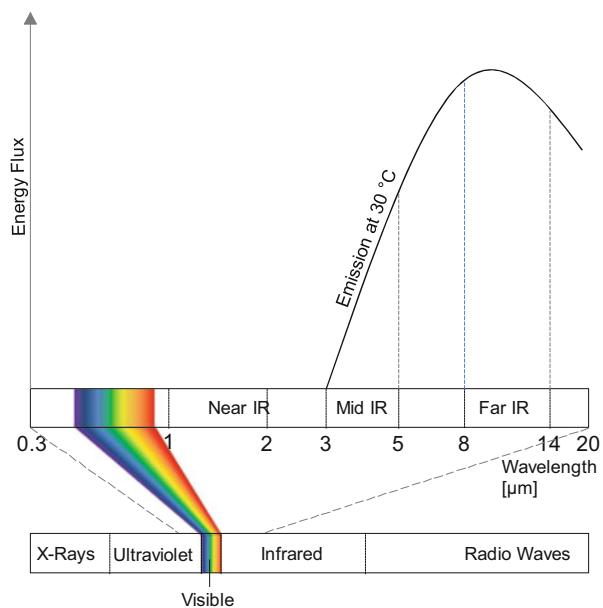
However, one significant disadvantage has to be set against these benefits:

- The emissivity of the object to be measured must be known and of sufficient magnitude.

This means that the temperature of transparent objects (e.g., semiconductors) or strongly reflecting objects (e.g., mirrors) cannot be determined.

The infrared spectral range comprises the wavelength range of the electromagnetic radiation spectrum from 780 nm to 1 mm (Fig. 1). The subdivision of this range is based on technical criteria (Table 1). Certain wavelength ranges of radiation, known as atmospheric windows, are hardly absorbed by the atmosphere. These are particularly suitable for thermography. A few basic relationships are shown in Table 2. The radiation of a body is displaced with rising temperature towards shorter wavelengths. Whereas a body at room temperature (ca. 30 °C) predominantly radiates in the far infrared, the thermal radiation of extremely hot objects takes place in the near infrared or visible spectral range. Visible thermal radiation sources include glowing steel (>800 °C) and the sun (surface temperature ca. 5800 K).

**Fig. 1** Electromagnetic radiation spectrum



**Table 1** The spectral ranges and important atmospheric windows; (a) general; (b) for thermography

a)	Range	Spectral range	Abbreviation	Wavelength range	Atmospheric windows
Optic	Visible	VIS		0.38–0.78 $\mu\text{m}$	0.4–0.7 $\mu\text{m}$
Infrared	Near infrared	NIR		0.78–2.0 $\mu\text{m}$	0.75–0.8 $\mu\text{m}$ 0.95–1.1 $\mu\text{m}$ 1.2–1.3 $\mu\text{m}$ 1.5–1.7 $\mu\text{m}$
	Mid infrared	MIR		2.0–6.0 $\mu\text{m}$	2.1–2.4 $\mu\text{m}$ 3.2–4.1 $\mu\text{m}$ 4.4–5.2 $\mu\text{m}$
	Far infrared	FIR		6.0–40 $\mu\text{m}$	8–13 $\mu\text{m}$
	Ultrafar infrared	UFIR		40–1000 $\mu\text{m}$	
b)	Infrared	Short wave	SW	0.78–1.7 $\mu\text{m}$	
		Mid wave	MW	3–5 $\mu\text{m}$	
		Long wave	LW	8–14 $\mu\text{m}$	

## Radiation Laws

The theoretical description of the emission of radiation of a body is based on a blackbody. A blackbody is an object, whose radiance is independent of the angle (Lambert radiator) and emits the maximum amount of energy possible at each

**Table 2** Relations of the fundamental constants;  $c$  speed of light,  $h$  Planck's constant

Name	Relation	Unit	Range NIR $\leftrightarrow$ UFIR
Wavelength	$\lambda = \frac{c}{\nu}$	$\mu\text{m}$	0.78–1000 $\mu\text{m}$
Wave number	$\sigma = \frac{1}{\lambda} = \frac{\nu}{c}$	$\text{cm}^{-1}$	12,000–10 $\text{cm}^{-1}$
Frequency	$\nu = \frac{c}{\lambda}$	$\text{s}^{-1}$ or Hz	$3.85 \cdot 10^{14}$ – $3 \cdot 10^{11}$ Hz 385 THz–300 GHz
Photon energy	$E = h\nu = \frac{hc}{\lambda}$	eV	1.6 eV–1.24 keV

wavelength. The spectral radiance  $L_{\lambda B}$  of a blackbody is described by Planck's Law of Radiation:

$$L_{\lambda B} = \frac{2hc^2}{\lambda^5 \Omega_0} \frac{1}{e^{\frac{hc}{k_B T}} - 1}, \quad (1)$$

where  $h$  is Planck's constant,  $c$  the speed of light,  $k_B$  the Boltzmann constant,  $T$  the absolute temperature, and  $\Omega_0$  the unit solid angle (steradian). The subscript  $B$  denotes that the relationship applies to a blackbody. For real radiators (bodies), it is necessary to multiply the equation by the wavelength-dependent emissivity  $\varepsilon_\lambda$  (refer to Kirchhoff's Law from Eq. (17)).

In order to calculate radiation fluxes, the spectral radiant exitance  $M_{\lambda B}$  is often required. This is the exitance given off into the half-space per surface area and per wavelength:

$$M_{\lambda B} = \pi \Omega_0 L_{\lambda B}. \quad (2)$$

From Eq. (1), the familiar form of Planck's Law of Radiation is obtained (Fig. 2a):

$$M_{\lambda B} = \frac{c_1}{\lambda^5} \frac{1}{e^{\frac{c_2}{\lambda T}} - 1} \quad (3)$$

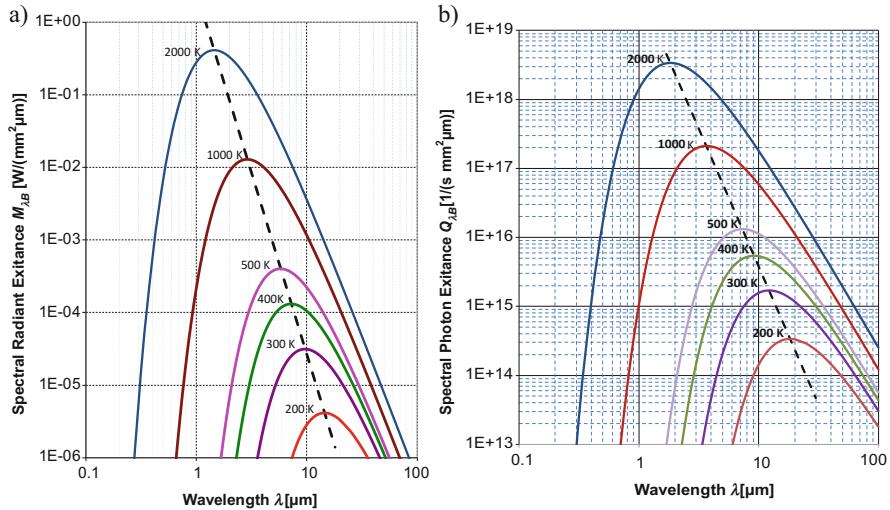
with the radiation constants

$$c_1 = 2\pi hc^2 = 3.741832 \cdot 10^{-16} \text{ W m}^2 \quad (4)$$

and

$$c_2 = \frac{hc}{k_B} = 1.438786 \cdot 10^{-2} \text{ K m}. \quad (5)$$

Also of interest are the number of photons radiated per unit time, surface area, and wavelength  $Q_{\lambda B}$ . By using the relationship for the energy of a photon as given by



**Fig. 2** Planck's Law of Radiation: (a) specific radiation and (b) specific photon radiation of a blackbody; parameter: temperature. The dashed lines depict Wien's Displacement Law from Eqs. (9) or (10)

$$W = h\nu = \frac{hc}{\lambda} \quad (6)$$

it is possible to calculate from Eq. (3) the spectral photon exitance  $Q_{\lambda B}$  (Fig. 2b):

$$Q_{\lambda B} = \frac{c'_1}{\lambda^4} \frac{1}{e^{\frac{c_2}{\lambda T}} - 1} \quad (7)$$

with the radiation constant

$$c'_1 = 2\pi c = 1.883634 \cdot 10^9 \text{ m s}^{-1}. \quad (8)$$

With increasing temperature, the maximum energy of the radiation is displaced towards lower wavelengths, whereby the product of wavelength and temperature remains constant. This relationship is described in Wien's Displacement Law:

$$\lambda_{max}T = 2898 \text{ } \mu\text{m K}. \quad (9)$$

The following applies to photon flux:

$$\lambda_{Q,max}T = 3669 \text{ } \mu\text{m K}. \quad (10)$$

At a temperature of, say,  $23^\circ\text{C}$ , the maximum radiation wavelength is  $\lambda_{max} = 9.8 \text{ } \mu\text{m}$  or  $\lambda_{Q,max} = 12.4 \text{ } \mu\text{m}$ .

The radiant exitance  $M_B$  in a particular wavelength range  $\lambda_1$  to  $\lambda_2$  is calculated by integrating over the corresponding wavelength range  $\lambda_1$  to  $\lambda_2$  from Planck's Law of Radiation:

$$M_B = \int_{\lambda_1}^{\lambda_2} M_{\lambda B} d\lambda. \quad (11)$$

Unfortunately, only one closed-form solution exists for this integral in the wavelength range from  $\lambda_1 = 0$  to  $\lambda_2 \rightarrow \infty$ , namely, the Stefan-Boltzmann Law:

$$M_B = \sigma T^4 \quad (12)$$

with the Stefan-Boltzmann Constant

$$\sigma = \frac{2\pi^5 k_B^4}{15 c^2 h^3} = 5.67 \cdot 10^{-8} \frac{\text{W}}{\text{m}^2 \text{K}^4}. \quad (13)$$

For the photon exitance  $Q_B$ ,

$$Q_B = \sigma' T^3 \quad (14)$$

equivalently applies, where the Stefan-Boltzmann constant for the photon rate is

$$\sigma' = \frac{4\pi\zeta(3)k_B^3}{h^3 c^3} = 1.5204 \cdot 10^{15} \text{s}^{-1} \text{m}^{-2} \text{K}^{-3} \quad (15)$$

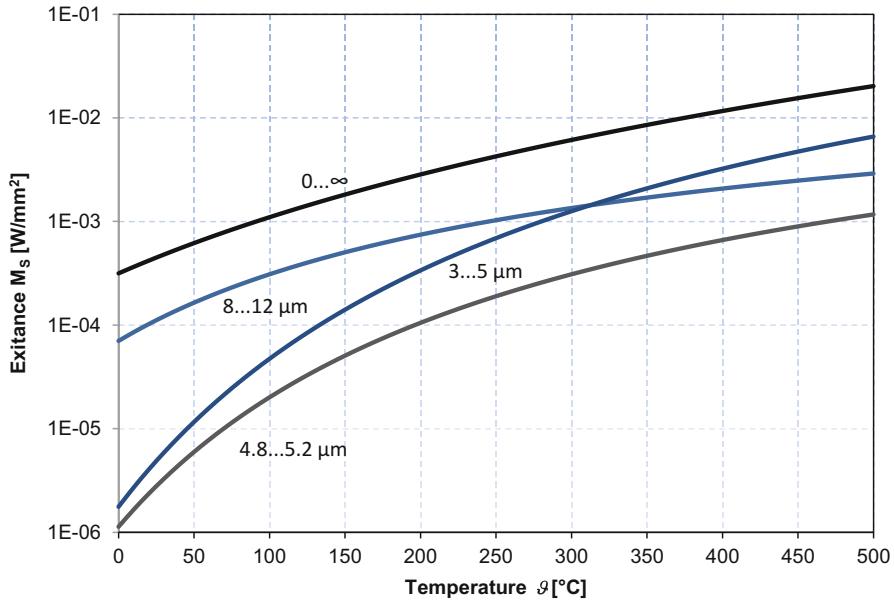
and the Riemann zeta function  $\zeta(3) = 1.202056\dots$ . The Stefan-Boltzmann Law describes the total radiation of a blackbody in the half-space. A blackbody at a temperature of 100 °C radiates, for example, 1087.1 W/m<sup>2</sup> or 7.8·10<sup>22</sup> photons/(s·m<sup>2</sup>) (Fig. 3).

## IR Characteristics of Bodies

Each body that emits radiation also absorbs radiation. This state of affairs is described by Kirchhoff's Law of Radiation. Under the assumption that a body is a Lambert radiator, it states that the spectral emissivity  $\varepsilon_\lambda$  is equal to the spectral absorptivity  $\alpha_\lambda$  as follows:

$$\varepsilon_\lambda = \alpha_\lambda. \quad (16)$$

As for blackbodies, the emissivity is a maximum ( $\varepsilon_\lambda = 1$ ), then all incident radiation is also absorbed ( $\alpha_\lambda = 1$ ). It therefore appears black. For real bodies, the following applies:



**Fig. 3** Blackbody radiation. Parameter: wavelength range

$$M_\lambda = \varepsilon_\lambda M_{\lambda B} \quad (17)$$

or

$$L_\lambda = \varepsilon_\lambda L_{\lambda B}. \quad (18)$$

The emissivity is the relation of the spectral specific radiation of a body to that of a blackbody for a given temperature  $T$  and wavelength  $\lambda$ :

$$\varepsilon(\lambda, T) = \frac{M_\lambda(\lambda, T)}{M_{\lambda B}(\lambda, T)}. \quad (19)$$

If a body is not a Lambert radiator, i.e., its radiance is direction-dependent, then the emissivity is calculated as follows:

$$\varepsilon(\lambda, T, \vartheta, \varphi) = \frac{L_\lambda(\lambda, T, \vartheta, \varphi)}{L_{\lambda B}(\lambda, T)}, \quad (20)$$

where  $\vartheta$  is the azimuth and  $\varphi$  the zenith angle of the emitted radiation. In many cases, the emissivity in a specific wavelength range  $\lambda_1$  to  $\lambda_2$ , the so-called band emissivity  $\varepsilon_B$ , is of interest:

$$\varepsilon_B(\lambda, T, \vartheta, \varphi) = \frac{\int_{\lambda_1}^{\lambda_2} L_\lambda(\lambda, T, \vartheta, \varphi) d\lambda}{\int_{\lambda_1}^{\lambda_2} L_{\lambda B}(\lambda, T) d\lambda}. \quad (21)$$

For the total radiation ( $\lambda_1 = 0$  to  $\lambda_2 \rightarrow \infty$ ), we obtain:

$$\varepsilon(\lambda, T, \vartheta, \varphi) = \frac{\int_0^\infty L_\lambda(\lambda, T, \vartheta, \varphi) d\lambda}{\int_0^\infty L_{\lambda B}(\lambda, T) d\lambda} = \frac{\int_0^\infty L_\lambda(\lambda, T, \vartheta, \varphi) d\lambda}{\frac{\sigma T^4}{\pi}}. \quad (22)$$

The radiation characteristics of a body is of enormous importance for calculating its temperature. It is only with a known emissivity that the temperature of the body can be determined with Planck's Formula from Eqs. (3) or (7). The emissivity of real bodies essentially depends on

- Its temperature
- Its surface structure and characteristics
- The wavelength
- The direction of the radiation emitted

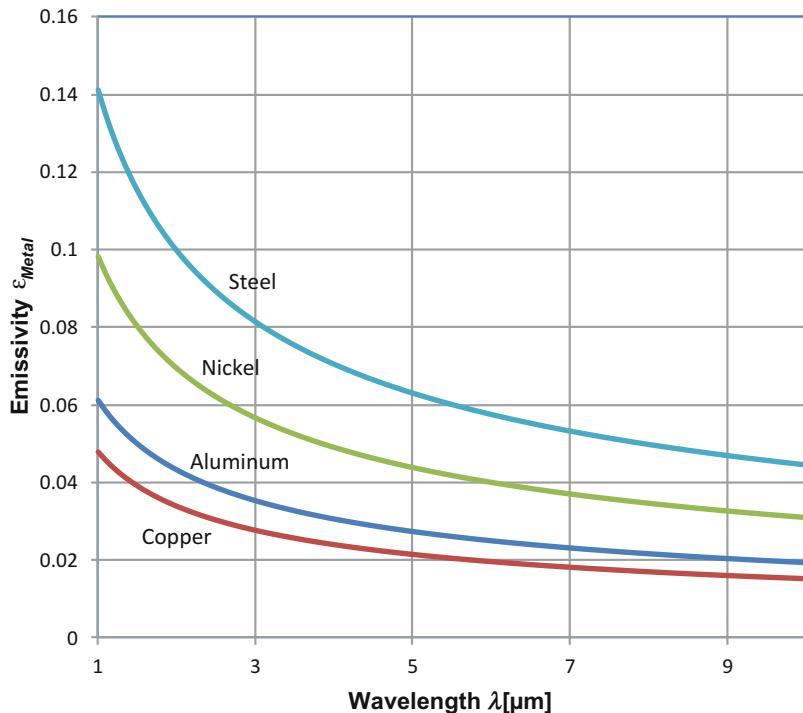
To date, theories for the calculation of  $\varepsilon$  are only known for ideal gases and metals with an ideal surface. All other emissivities are based on empirically determined relationships or – as is mostly the case – on measurements. By systematising the emission characteristics of real materials, it turns out that specific material categories, such as

- Metals
- Dielectric materials
- Oxides
- Ionic crystals
- Glass
- Organic substances
- Biological materials

possess similar emissivity levels.

For metals with the DC conductivity  $\kappa_{\text{Metal}}$ , the following equation applies to the emissivity measured perpendicular to the surface (Hagen-Rubens Relation, Fig. 4):

$$\varepsilon_{\text{Metal}}(\lambda) = 4 \sqrt{\frac{\pi c_0 \varepsilon_0}{\kappa_{\text{Metal}} \lambda}} \quad (23)$$



**Fig. 4** Emissivity of metals based on the Hagen-Rubens relation from Eq (23)

or

$$\varepsilon_{\text{Metal}}(\lambda) = \frac{0.365}{\sqrt{\kappa_{\text{Metal}}}\lambda}, \quad (24)$$

where  $c_0$  is the speed of light in a vacuum and  $\varepsilon_0$  the dielectric constant of the vacuum. Since the DC conductivity  $\kappa_{\text{metal}}$  is approximately indirectly proportional to the temperature, then the following applies:

$$\varepsilon_{\text{Metal}} \sim \sqrt{\frac{T}{\lambda}}. \quad (25)$$

Unfortunately, the relation expressed in Eq. (24) only applies to ideal metal surfaces. Oxide layers or surface irregularities exhibit substantially deviating values for the emissivity. Tables 3 and 4 depict a few real emissivities.

Radiation can not only be absorbed and emitted by a real body, however, but also penetrate (with spectral transmissivity  $\tau_\lambda$ ) and reflected (with spectral reflectivity  $\rho_\lambda$ ). In terms of energy, the following applies:

**Table 3** Emissivities of various nonmetals /[www.raytek.com/](http://www.raytek.com/)/

Wavelength	5.0 $\mu\text{m}$	7.9 $\mu\text{m}$	8–14 $\mu\text{m}$
Asphalt	0.9	0.95	0.95
Basalt	0.7	0.7	0.7
Concrete	0.9	0.95	0.95
Ice		0.98	0.98
Earth	0.9–0.98	0.9–0.98	
Glass plate	0.98	0.85	0.85
Wood	0.9–0.95	0.9–0.95	0.9–0.95
Ceramics	0.85–0.95	0.95	0.95
Paper (any color)	0.95	0.95	0.95
Cloth	0.95	0.95	0.95
Sand	0.9	0.9	0.9
Snow		0.9	0.9
Water		0.93	0.93

**Table 4** Emissivities of various metals /[www.raytek.com/](http://www.raytek.com/)/

Wavelength	1.0 $\mu\text{m}$	1.6 $\mu\text{m}$	8–14 $\mu\text{m}$
Aluminum unoxidized	0.1–0.2	0.02–0.2	
Aluminum oxidized	0.4	0.4	0.2–0.4
Iron	0.35	0.1–0.3	
Iron oxidized	0.4–0.8	0.5–0.9	0.5–0.9
Gold	0.3	0.01–0.1	
Copper oxidized	0.4–0.8	0.2–0.9	
Steel cold-rolled	0.8–0.9	0.8–0.9	0.7–0.9
Steel blank	0.35	0.2–0.9	0.1–0.8

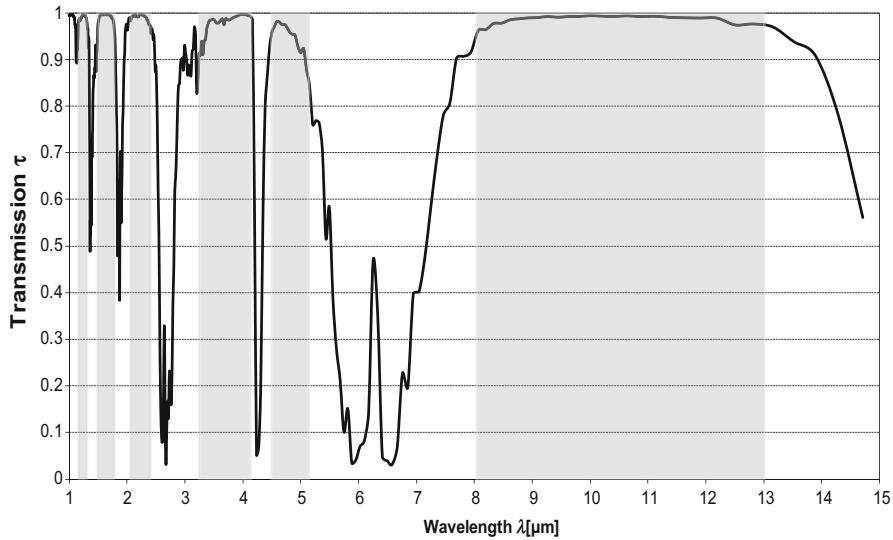
$$\alpha_\lambda + \tau_\lambda + \rho_\lambda = 1. \quad (26)$$

If the absorption, emission, transmission, and reflection characteristics are independent of the wavelength, the bodies are referred to as grey radiators. From Eq. (26), it then follows:

$$\alpha + \tau + \rho = 1. \quad (27)$$

There exist bodies, which exhibit no absorption and thus no emission, such as ideal mirrors ( $\rho = 1$ ) and ideal windows ( $\tau = 1$ ). In these cases, no determination of the temperature of the bodies using infrared radiation is possible.

A decisive factor in the use of noncontact temperature measurement techniques is the transmission of the measurement path. This is primarily determined by the atmospheric transmission of IR radiation (Fig. 5). The gases present in the atmosphere absorb electromagnetic radiation in so-called absorption bands. Particularly noteworthy in this regard are carbon dioxide



**Fig. 5** Transmission of the atmosphere for a measurement path of 10 m. Atmospheric windows are marked in grey

(strong absorption bands at 2.7; 4.3 and 14.7  $\mu\text{m}$ ) and water vapor (at 1.4; 1.8; 2.7; 5.8 and 6.7  $\mu\text{m}$ ). The influence of the atmosphere can be kept to a minimum by the use of atmospheric windows with a high transmission of radiation (Table 1).

If a body is not transparent ( $\tau_O = 0$ ) and its emissivity  $\varepsilon_O < 1$ , then from Eq. (27) there follows for its emissivity

$$\rho_O = 1 - \varepsilon_O. \quad (28)$$

From Eq. (24), the radiance emitted is defined as

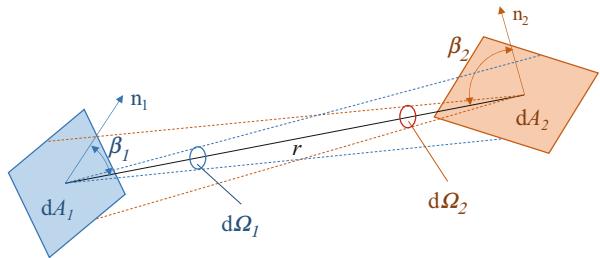
$$L \sim \tau_A [\varepsilon_O L_O(T_O) + (1 - \varepsilon_O)L(T_U)]. \quad (29)$$

Here all variables with the subscript  $O$  describe the object measured,  $\tau_A$  is the transmission of the atmosphere, and  $T_U$  the ambient temperature of the object. Eq. (29) expresses that the radiance of a real body is not only dependent on its temperature but also on the radiation reflected by the body and thus on its environment.

## Basic Law of Photometry

The Basic Law of Photometry describes the radiation flux  $\Phi_{12}$  reaching a cross-section  $dA_2$  from a cross-section  $dA_1$  as:

**Fig. 6** Definition of the geometric parameters for the Basic Law of Photometry.  $n_1$  and  $n_2$  are the surface normals of the cross-sections  $dA_1$  and  $dA_2$



$$d^2\Phi_{12} = L_1 \frac{dA_1 \cos \beta_1 dA_2 \cos \beta_2}{r^2}. \quad (30)$$

The distance  $r$  represents the shortest connection of the two cross-sections  $dA_1$  and  $dA_2$ . The terms  $dA_1 \cos \beta_1$  and  $dA_2 \cos \beta_2$  are the projections of the cross-sections at right angles to the distance vector  $r$  (Fig. 6).

Since each cross-section both emits and receives radiation, the radiation flux exchanged  $\Phi$  represents the radiation balance of the cross-sections, as follows:

$$d^2\Phi = d^2\Phi_{12} - d^2\Phi_{21} = \frac{dA_1 \cos \beta_1 dA_2 \cos \beta_2}{r^2} (L_1 - L_2). \quad (31)$$

A commonly used solution of Eq. (30) is the irradiance  $E_P$  of a point (cross-section) by a Lambert radiator with the radiance  $L$ :

$$E_P = \pi L \sin^2 \varphi \Omega_0 \quad (32)$$

where  $\varphi$  ( $= FOV/2$ ) is the opening angle of the irradiated cross-section.

## Sensors

In order to understand the functionality of thermographic systems, basic knowledge of the infrared radiation detectors used in these systems is required. Currently, only matrix or imaging sensors, the so-called FPAs (focal plane arrays), are used in thermography systems (older systems use scanners). The detection of infrared radiation is based on two physical operating principles. These are the measurement of the radiation energy or radiant power using thermal sensors (Budzier and Gerlach 2011) and the recording of the photon flux using photon sensors (Rogalski 1995). Before we present these two types of sensors, though, it is necessary to explain a few important sensor characteristics.

## Sensor Characteristics

The sensitivity is defined as the relation of the change in the output signal, in this case the change in output voltage  $\Delta U_S$ , to the change in an input signal – for infrared

radiation sensors, this is the change in radiation flux  $\Delta\Phi_O$ . For the so-called voltage sensitivity (or responsivity)  $R_V$  we therefore obtain:

$$R_V = \frac{\Delta U_S}{\Delta\Phi_O}. \quad (33)$$

The signal-to-noise ratio of a sensor is described by the *NEP* (noise equivalent power):

$$NEP = \frac{\tilde{u}_R}{\Delta U_S} \Delta\Phi_O = \frac{\tilde{u}_R}{R_V}, \quad (34)$$

where  $\tilde{u}_R$  is the effective noise voltage. The *NEP* corresponds to the radiation flux, which creates a signal-to-noise ratio of one. The detectivity  $D$  is commonly specified in place of the *NEP*, as follows:

$$D = \frac{1}{NEP}. \quad (35)$$

Greater detectivity values also indicate qualitatively better sensors. For a comparison of sensors, the specific detectivity  $D^*$  is often used, which performs a normalization across the sensor surface  $A_S$  and bandwidth  $B$ :

$$D^* = \frac{\sqrt{A_S}}{\tilde{u}_{Rn}} R_V = \frac{\sqrt{A_S B}}{NEP}. \quad (36)$$

$\tilde{u}_{Rn}$  is here the noise voltage normalized  $\sqrt{B}$  across. The absolute detectability threshold is described by the BLIP detectivity  $D^*\text{BLIP}$  (BLIP: background-limited infrared performance). This occurs when the radiation noise remains the sole non-compensatable source of noise due to background radiation (refer to Fig. 19).

The temperature resolution is described by the noise-equivalent temperature difference *NETD*:

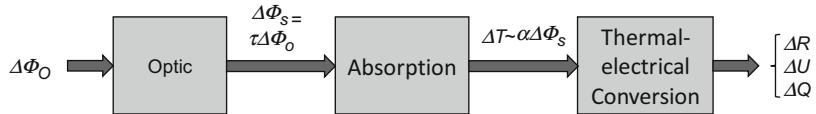
$$NETD = \frac{\Delta T}{\frac{\Delta U_S}{\tilde{u}_R}}. \quad (37)$$

The *NETD* is that temperature difference in the object, which generates a signal-to-noise ratio of one in the sensor.

As is common with electro-optical systems, the spatial resolution is described by the modulation transfer function *MTF*.

## Thermal Sensors

Part of the radiation emitted by an object  $\Delta\Phi_O$  is mapped by an optical system exhibiting the transmission  $\tau$  onto a sensor element, being absorbed by it (Fig. 7).



**Fig. 7** Operating principle of thermal sensors

This results in a change in temperature in the sensor element  $\Delta T$ , which, in turn, causes a change in an electrical quantity. Depending on the type of sensor, this can be a change in resistance ( $\Delta R$ ; bolometer), charge ( $\Delta Q$ ; pyroelectric sensor), or a voltage or current ( $\Delta U, \Delta I$ ; thermo-electrical sensor).

The sensor element consists of a thin, thermally well-insulated plate with a heat capacity  $C_p$  and is connected to the environment via a heat conduction coefficient  $G_{th}$  (Fig. 8). This simple thermal model can be used to easily calculate the change in temperature  $\Delta T$  of the sensor element for harmonic events with the frequency  $\omega = 2\pi f$ :

$$|\Delta T| = \frac{\alpha\tau}{G_{th}} \frac{|\Delta\Phi|}{\sqrt{1 + \omega^2\tau_{th}^2}}. \quad (38)$$

Here the frequency  $f$  denotes the modulation frequency of the radiation flux – not the frequency  $\nu$  of the radiation as specified in Table 2. For the thermal time constant  $\tau_{th}$  in Eq. (33), we obtain:

$$\tau_{th} = \frac{C_{th}}{G_{th}}. \quad (39)$$

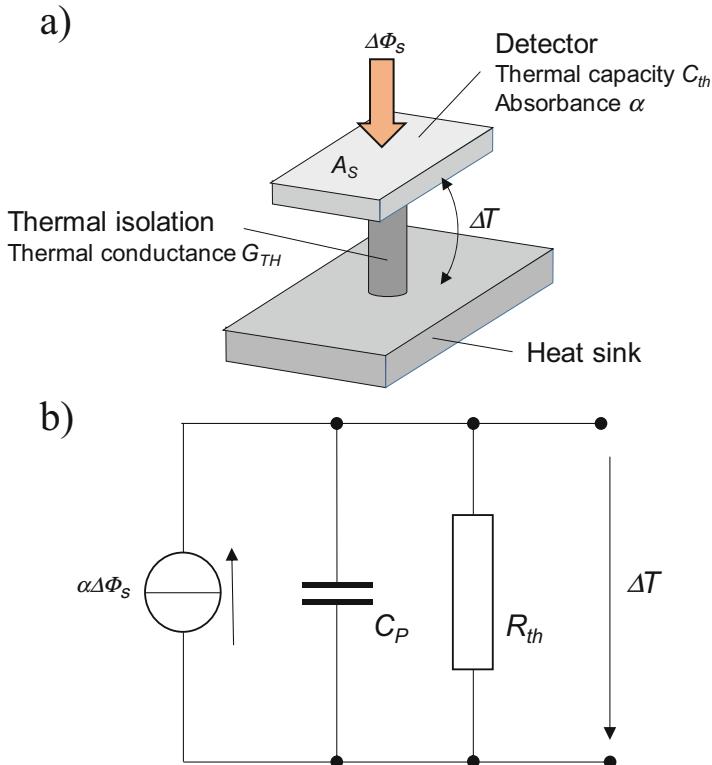
The sensitivity of thermal infrared sensors accordingly behaves in relation to the change over time of the object temperature as a first-order low-pass with the time constant  $\tau_{th}$ .

For the heat capacity of a cuboid sensor element, the following applies:

$$C_{th} = c_S \rho_S A_S d_S. \quad (40)$$

The heat capacity is thus dependent on the specific thermal capacity  $c_S$ , the density  $\rho_S$  of the sensor material, the sensor surface  $A_S$ , and the thickness  $d_S$  of the sensor. The heat exchange between the sensor element and the environment (heat sink) ensues via the thermal radiation, via the heat transfer by the surrounding gas as well as via the mechanical mounting of the sensor element. With most sensors, the last-named is responsible for the major part of the heat exchange.

In order to achieve the greatest possible temperature difference in the sensor element – and thus the greatest possible sensitivity, the heat capacity  $C_{th}$  – and therefore the thickness  $d_S$  – must be as small as possible, whereas the thermal resistance  $1/G_{th}$  – and thus the thermal isolation – must be as great as possible.



**Fig. 8** Thermophysical model of a thermal infrared sensor

The maximum detectivity is attained when the heat output from the sensor element solely takes place in the form of radiation. This is ensured by operating the sensor element in a vacuum (no convection and no heat transfer by the surrounding gas) and a thermally well-insulating mounting (the thermal conductivity value from the mounting must be much smaller than from the radiation). The maximum theoretically possible specific detectivity is thus:

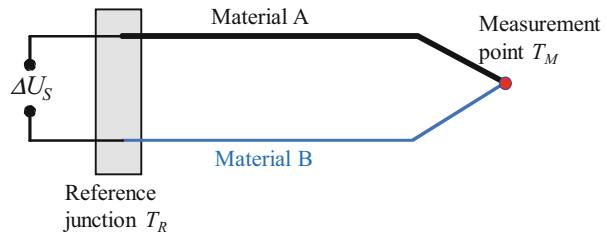
$$D_{BLIP}^* = \frac{1}{\sqrt{16k_B\sigma T^5}}. \quad (41)$$

For a thermal sensor at room temperature ( $T = 300$  K), the maximum specific detectivity is  $1.8 \cdot 10^{10}$  cm·Hz $^{1/2}$  W $^{-1}$ .

### Radiation Thermocouples

The operating principle of radiation thermocouples is based on the Seebeck effect. If two conductors with different Seebeck coefficients  $\alpha_{SA}$  and  $\alpha_{SB}$  are connected and the connection points are brought to different temperatures  $T_R$  and  $T_M$ , then a

**Fig. 9** Principle of a radiation thermocouple



thermoelectric voltage can be measured at the ends of the conductors (Fig. 9). As the Seebeck coefficients are extremely small, several of such thermocouples are mostly wired in series (typically  $N = 40 \dots 150$ ). These arrangements are referred to as thermopiles.

For the signal voltage of a thermopile, the following applies:

$$\Delta U_S = N(\alpha_{SM} - \alpha_{SR})(T_R - T_M) = N\Delta\alpha_S\Delta T \quad (42)$$

The specific detectivity is:

$$D^* = R_V \sqrt{\frac{A_S}{4k_B T R_T}}. \quad (43)$$

Typical specific detectivities for thermopiles are approx.  $2 \cdot 10^8 \text{ cm} \cdot \text{Hz}^{1/2} \text{ W}^{-1}$  (Poly-Si/Al) and ca.  $8 \cdot 10^8 \text{ cm} \cdot \text{Hz}^{1/2} \text{ W}^{-1}$  (Bi/Sb).

### Pyroelectric Sensors

Pyroelectric sensors use the pyroelectric effect for the thermoelectric conversion. The sensitivity of a pyroelectric sensor is given by

$$R_V = \frac{\alpha \pi_P A_S R}{G_{th}} \frac{\omega}{\sqrt{1 + \omega^2 \tau_{th}^2} \sqrt{1 + \omega^2 \tau_{el}^2}}. \quad (44)$$

where  $\tau_{el}$  is the electronic time constant and  $\pi_P$  the pyroelectric coefficient. For static signals, the sensitivity is therefore zero (Fig. 10). Pyroelectric sensors are not sensitive to d.c. light!

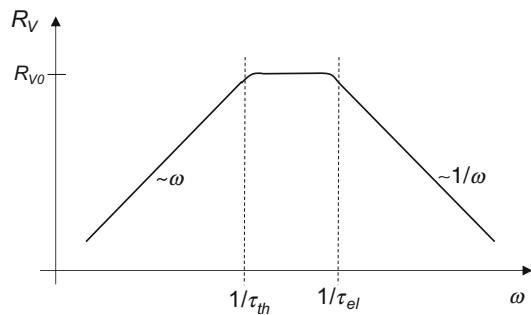
In the pyroelectric sensor, there exists not only radiation noise but also a series of other sources of noise. The most important of these are the  $\tan\delta$  noise caused by capacity losses as well as the current and voltage noises of the preamplifier. The specific detectivity strongly depends on the pyroelectric material used. The best values are obtained at a frequency of 10 Hz at approx.  $8 \cdot 10^9 \text{ cm} \cdot \text{Hz}^{1/2} \text{ W}^{-1}$ .

### Bolometers

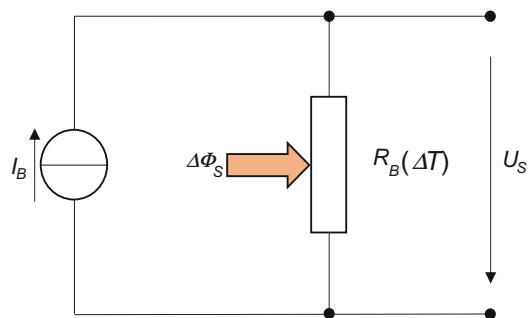
Bolometers make use of the temperature dependence of a resistor  $R_B$  for the thermoelectric conversion (Fig. 11).

The specific detectivity of a bolometer is given by:

**Fig. 10** Principal curve of the sensitivity of a pyroelectric sensor



**Fig. 11** Principle of a bolometer



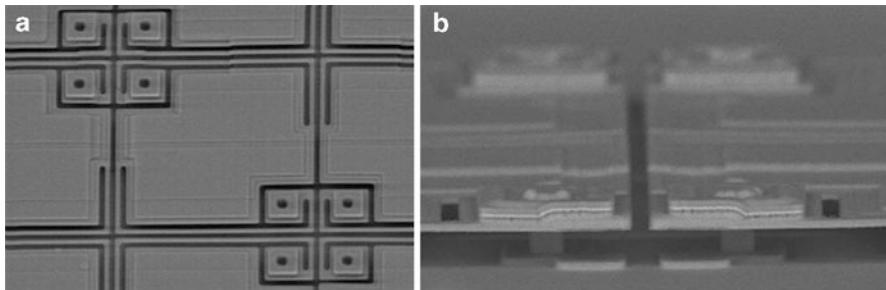
$$D^* = \frac{\alpha\tau\beta I_B}{G_{th}\sqrt{1 + \omega^2\tau_{th}^2}} \sqrt{\frac{A_S R_B}{4k_B T}} \quad (45)$$

Metal and semiconductor resistors are used as construction materials for a bolometer. While metal resistors possess a positive temperature coefficient, for semiconductor resistors it is negative. When using metal resistors, the sensitivity is typically 30 V/W, whereas it is approx.  $2 \cdot 10^4$  V/W when using semiconductor resistors.

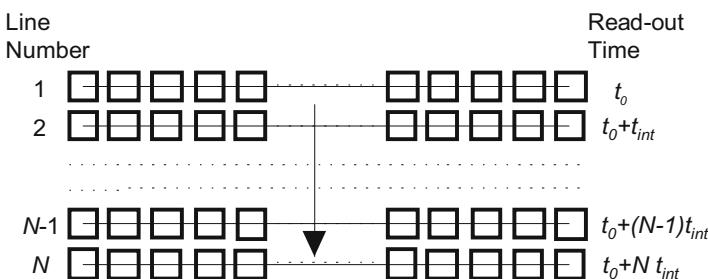
### Thermal Infrared Imaging Sensors

Micro-bolometer arrays are primarily used in thermographic cameras. In order to achieve the maximum possible thermal insulation of the individual pixels, these are realized as micro-bridges in a vacuum (Fig. 12). The micro-bridges are directly manufactured using a CMOS readout circuit. The height of the bridge is approx. 2.5 μm and is measured so that the bridge structure acts as a  $\lambda/4$  resonator to absorb IR radiation between 8 and 14 μm. Amorphous silicon (a-Si) and vanadium oxide ( $VO_x$ ) are used as construction materials for bolometers. The typical temperature resolutions (NETD) are less than 60 mK.

For each pixel, the readout circuit contains a preamplifier, a line and column multiplexer as well as the analogue control electronics required for the pixels. The pixel readout takes place line by line (rolling frame, Fig. 13). For this purpose, when



**Fig. 12** Photography of micro-bolometer bridges in a  $25 \mu\text{m}$  grid; (a) top view and (b) cross-section (Fraunhofer Institute for Microelectronic Circuits and Systems, Duisburg/Germany)



**Fig. 13** Readout principle of a micro-bolometer matrix with  $N$  lines and an image time of  $t_B = 1/f_B = Nt_{int}$

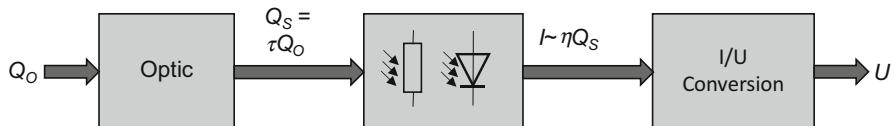
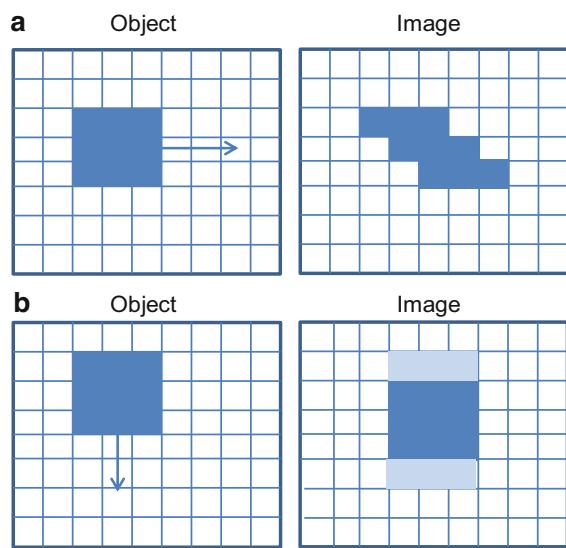
line  $n$  is read, line  $n + 1$  is provided with an electrical impulse. This thereby maximizes the length of the electrical impulse, the so-called integration time  $t_{int}$ . This enables the readout of the lines to ensue across the entire image time  $t_B$ . This is why smearing occurs with moving objects (Fig. 14). Typical image frequencies  $f_B$  are in the range of 50 Hz to 120 Hz. The maximum image frequency is limited by the thermal time constant (typically 5 to 10 ms).

## Photon Sensors

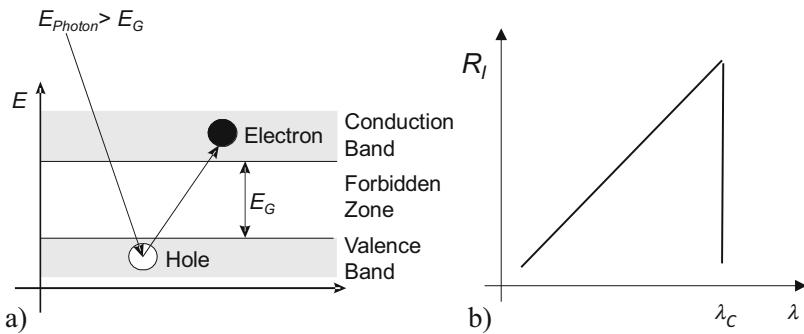
Photon sensors detect the incident radiation flux as a stream of photons and convert it into an electrical current (Fig. 15). They use the internal photoelectric effect to create electron-hole pairs by optically stimulating electrons from the valence band into the conduction band. Photon sensors are photon counters.

A photon striking the sensor element has a certain probability (quantum efficiency or quantum yield  $\eta$ ) of creating an electron-hole pair (Fig. 16a). The electrons created in the conduction band then result in the sensor current. The current sensitivity of a photon sensor is given by:

**Fig. 14** Smearing of moving objects: (a) horizontal motion; (b) vertical motion



**Fig. 15** Operational principle of photon sensors



**Fig. 16** Photon sensors: (a) generation of charge carriers in a semiconductor and (b) sensitivity of photon sensors as a function of wavelength

$$R_I = \frac{I}{Q_S} = \frac{\eta e}{hc} \lambda, \quad (46)$$

where  $e$  is the elementary charge,  $h$  is Planck's constant,  $c$  is the speed of light, and  $\lambda$  is the wavelength.

For the incident photons to be able to generate charges in the first place, the energy  $E$  of the photons must be greater than the band gap  $E_G$ :

$$E = h\nu = \frac{hc}{\lambda} > E_G. \quad (47)$$

This means that the sensitivity of photon sensors is proportional to the wavelength up to a cut-off wavelength

$$\lambda_C = \frac{hc}{E_C} \quad (48)$$

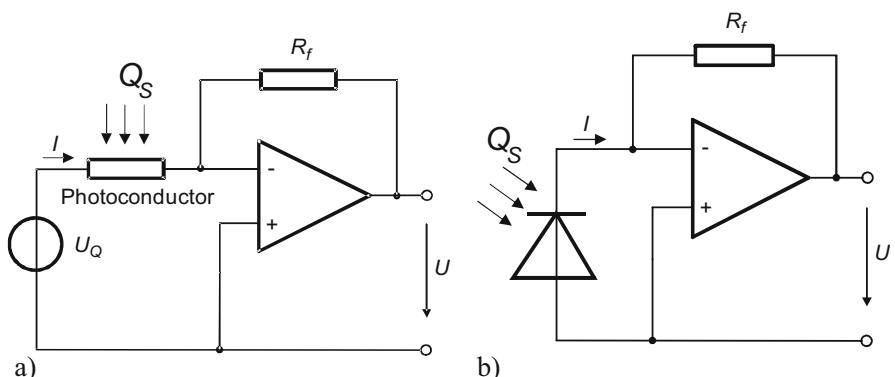
and – above this cut-off – then becomes zero (Fig. 16b). As the energy of infrared radiation is extremely low, it is necessary to use semiconductors with a small band gap. In order to prevent thermally stimulated charge carriers from flooding the conduction band and blocking the detection of the photocurrent, it is necessary to cool down photons for the mid and far infrared range. For the far infrared range, the working temperatures required of, say, photodiodes are 80 K and for, say, quantum mechanical sensors 60 K.

Semiconductor sensors are primarily used for detecting infrared radiation. These are divided into photoconductors (photoresistors), photodiodes, and quantum-well sensors.

### Photoconductors

In semiconductors, the conductivity changes depending on the number of charge carriers created in the conduction band. Photoconductors are thus semiconductor resistors, whose resistance is changed by the incident infrared radiation. This is the origin of the term photoresistor. Fig. 17a depicts the basic circuit of a photoresistor.

The maximum detectivity of a photoresistor is given by:



**Fig. 17** Basic circuit of (a) a photoconductor and (b) a photodiode in element manufacturing operations

$$D_{BLIP}^* = \frac{\lambda}{2hc} \sqrt{\frac{\eta}{Q_B}}. \quad (49)$$

where  $Q_B$  is the background photon flux.

Preferred materials for photoresistors are the intrinsic semiconductors silicon (Si), lead sulphide (PbS), lead selenide (PbSe), indium antimonide (InSb), and mercury cadmium telluride (CdHgTe, MCT). Also used are extrinsic (or impure) semiconductors, such as doped silicon (Si) and doped germanium (Ge).

### Photodiodes

If charge carriers are created in the depletion region of an electrically biased semiconductor pn junction by radiation, then these are known as photodiodes. If the diode voltage is poled in the blocking direction, then a saturation current flows through the diode. The photodiode is then – as with the photoresistor – in photoconducting mode. If the photodiode is operating without bias, say, in short-circuit mode or – more rarely – in idle mode, the photodiode is in element manufacturing mode and is referred to as a photo element. It is then working photovoltaically.

The maximum detectivity is a factor of  $\sqrt{2}$  higher than is the case with photoresistors, as follows:

$$D_{BLIP}^* = \frac{\lambda}{hc} \sqrt{\frac{\eta}{2Q_B}}. \quad (50)$$

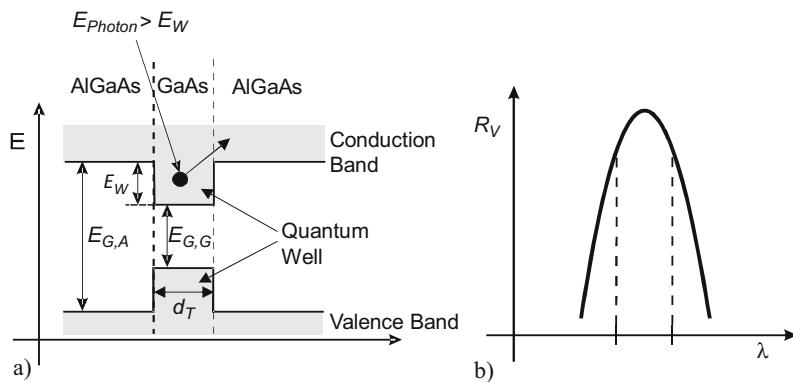
Frequently used semiconductors used for infrared photodiodes are indium antimonide (InSb) and mercury cadmium telluride (MCT).

A special form of photodiodes are the Schottky barrier diodes, consisting of a metal-semiconductor junction. Here platinum silicide (PtSi) is primarily used. This, however, possesses the disadvantage both of an extremely low quantum efficiency (<10%) and a maximum cut-off wavelength of 5  $\mu\text{m}$ . Notwithstanding this, its high characteristics reproducibility (homogeneity) for matrix sensors with a large number of pixels is a beneficial feature.

### Quantum-Well Sensors

If sensors make use of quantum-mechanical effects, such as the displacement of energy levels, to detect radiation, they are referred to as quantum-well sensors. Optoelectronic sensors primarily exploit the spatial restriction on the freedom of movement of charge carriers and the tunneling through potential barriers (quantum tunnelling effect) (Fig. 18a).

The basic principle is depicted by the quantum well (Fig. 18). If a thin layer ( $d_T < 30 \text{ nm}$ ) with a smaller band gap, say, GaAs with the band gap  $E_{G,G} < E_{G,A}$ , is embedded in a semiconductor, say, AlGaAs with a band gap of  $E_{G,A}$ , a spatially restricted potential lowering, the quantum well, is created. Charge carriers (electrons) that can no longer reach the conduction band are collected there. If a photon now hits a captured electron, the electron is able to regain the conduction band,



**Fig. 18** (a) Principle of a quantum well, and (b) the spectral sensitivity of a QWIP

thereby contributing to the conduction. As with the photoconductor, it is then possible to determine the absorbed photon stream from the conductivity (Fig. 17a).

Sensors that work on this principle are called quantum well infrared photo detector (QWIPs). To increase the quantum efficiency, several quantum wells are laid out next to each other. These groupings are known as multiple quantum wells (MQW). If the distance between the individual quantum wells is small, a lattice structure is created, in which the charge carriers can tunnel through the gaps between the individual quantum wells. This kind of structure is known as a superlattice. Superlattices allow the realization of photodiodes.

The characteristic wavelength dependence of the sensitivity is depicted in Fig. 18b. QWIP detectors possess a narrow frequency band (<20% bandwidth). The materials used are III-V compound semiconductors, such as the aforementioned semiconductors gallium arsenide (GaAs) and aluminum gallium arsenide (AlGaAs).

The BLIP detectivity corresponds to that from photoconductive and photovoltaic quantum sensors. It should be noted, however, that with  $\eta < 20\%$ , the quantum efficiency is significantly lower than with the photoresistors and photodiodes ( $\eta > 70\%$ ) described above. In order to achieve the same detectivities in practice, QWIP detectors are more strongly cooled, say, down to between 50 and 60 K.

The manufacturing practice involved (epitaxial growth of layers) enables several QWIP or superlattice detectors to be laid out on top of each other. This allows detectors with sensitivities in different wavelength ranges (dualband sensors, multispectral sensors) to be readily manufactured.

### Photon Imaging Sensors

IR photon sensors are generally hybrid designs, consisting of a silicon-based CMOS readout circuit and the sensor chip. This enables a maximum utilization of the sensor surface while allowing the partly CMOS-noncompatible manufacturing methods of photon sensors. As readout technologies, both a full image readout (i.e., the simultaneous scanning of all pixels – the snapshot mode) and a line-by-line readout

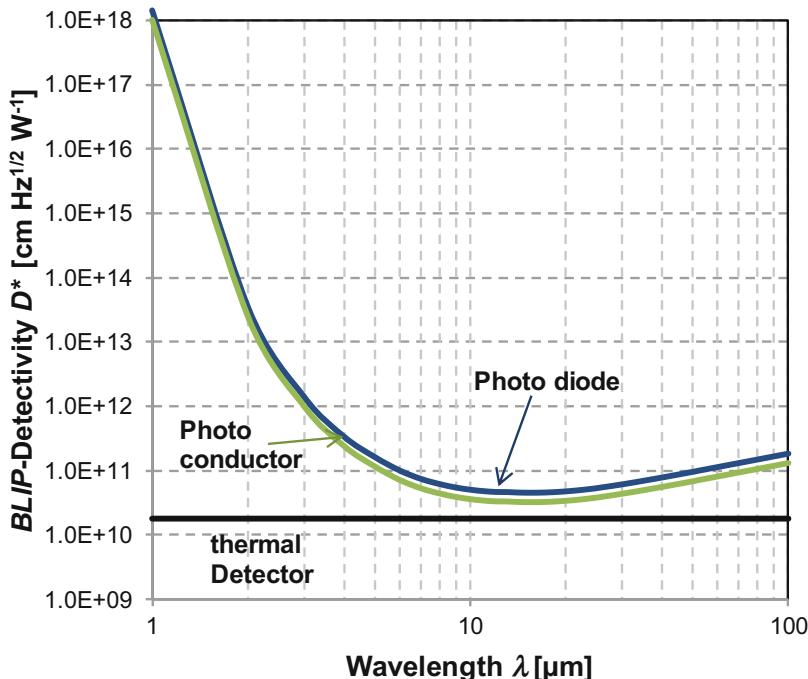
(rolling frame, refer to Fig. 16) are used. The maximum image frequency is only determined by the readout circuit and – in line with the television standard – is usually in the range of 50 Hz to 200 Hz. The readout circuits generally contain a variety of processing options. For example, part-images with a significantly higher image frequency can be read out.

Particularly interesting in this regard are the possibilities of manufacturing dualband matrix sensors, which are also known as colour sensors.

### Comparison of Thermal and Photon Sensors

If a comparison is made of thermal sensors and photon sensors, then photon sensors have the following advantages and disadvantages vis-à-vis thermal sensors:

- The sensitivity and the detectivity are dependent on the wavelength.
- The sensitivity and the detectivity are considerably greater: By up to several powers of magnitude – Depending on the wavelength (Fig. 19).
- The time constants are far lower:  $\tau_{PH} = 0.1 \dots 0.001 \tau_{th}$ .



**Fig. 19** BLIP detectivity of thermal sensors, photoconductors, and photodiodes. Background temperature 300 K;  $FOV = 2\pi$

- Photon sensors have to be kept at temperatures significantly below room temperature.
  - The costs are very much higher.
- 

## Thermal Imaging Cameras

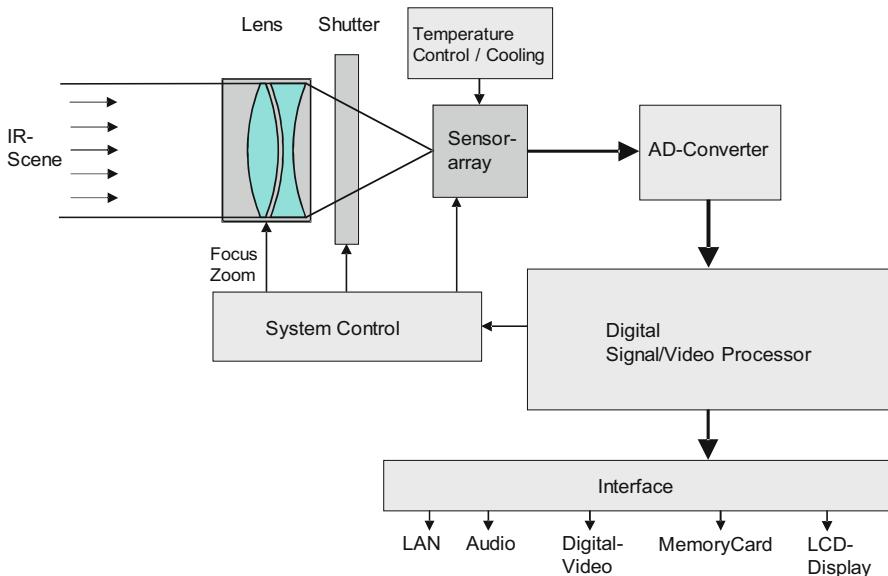
Thermal imaging cameras (thermographic cameras, infrared cameras) are imaging systems, which provide a first approximation of the temperature distribution of a scene (Williams 2009). They present the radiation emitted by a body or object in a two-dimensional image (thermal image, thermogram, infrared image). The radiation emitted from an object is composed of the intrinsic radiation, the reflected ambient radiation, and possibly the transmitted background radiation (refer to section “[IR Characteristics of Bodies](#)”). For this reason, a thermal image is only an approximate depiction of the temperature distribution in the scene.

In the atmospheric windows, thermal imaging cameras function in the mid infrared ( $\lambda = 3$  to  $5 \mu\text{m}$ ) or the far infrared ( $\lambda = 8$  to  $13 \mu\text{m}$ ). For short measuring distances – such as occur in nondestructive testing, which spectral range is favored basically depends on the radiation characteristics (emissivity and temperature) of the object measured. Modern thermal imaging cameras often include a second camera in the visible spectral range. This, however, merely serves for better visualization of the measurement results – not for measuring the temperature.

According to the sensor used, the terms “cooled systems” or “uncooled systems” are used. Cooled thermal imaging cameras use photon sensors and therefore always include an appropriate cooling system. Due to the high sensitivity of photon sensors, cooled systems are also suitable for very narrowband and multispectral applications. Uncooled thermal imaging cameras make use of thermal sensors (usually microbolometers). The sensors are often kept at a particular temperature using a Peltier element, in order to stabilize the sensor temperature close to room temperature. As a result, they possess a significantly lower power consumption and smaller dimensions than cooled cameras. Due to their relatively low sensitivity in the mid infrared range, they are only used for measurements at higher temperatures ( $>200^\circ\text{C}$ ).

The basic design of thermal imaging cameras corresponds to that of conventional video cameras. For this reason, only a few selected components are described in detail below.

Figure 20 shows the design of a thermal imaging camera. An infrared optical system displays the measuring scene onto the imaging sensor. The lenses of this system consist of germanium, silicon or special glass. For uncooled systems, wide-opening optics ( $f\text{-number } k < 1.4$ ) are used. An important feature of uncooled thermal cameras is the mechanical shutter. It is cyclically closed to recalibrate the system, whereby the cycle time is dependent on the temperature change of the camera, particularly the optical channel, and is usually several minutes. The shutter operates for a few hundred milliseconds. As this can be very distracting during the measuring process, it is often possible to trigger the shutter. This allows the shutter to



**Fig. 20** The design of a thermal imaging camera

operate for the optical channel at a defined and – as far as possible – less disturbing time, and the system to be recalibrated. For cooled systems, either the shutter is dispensed with or the shutter cycle time is set to several hours.

The digitalization (AD conversion) of the sensor signal ensues on a 14 to 16 bit basis. The image data is then processed in a signal processor. For outputting the data, the usual interfaces are generally used. For industrial measurement technology, this is primarily CameraLink and Ethernet.

For documenting and analyzing the measurements, modern portable thermogram systems possess not only the aforementioned visual camera but also an audio interface for recording comments and a laser pointer for marking the measuring points.

The most important parameters of a thermal imaging camera are the thermal and spatial resolution. All figures always refer to blackbodies.

The thermal resolution of a thermal imaging camera is described by the relative temperature resolution *NETD* and an absolute measurement error. The *NETD* describes the smallest measurable temperature difference in a scene. It is generally calculated as follows:

$$NETD = \frac{4k^2 + 1}{\sqrt{A_P \tau_{opt}}} \frac{\sqrt{B}}{\varepsilon \int_{\lambda_1}^{\lambda_2} D^*(\lambda) \frac{dM_{\lambda B}}{dT} d\lambda}. \quad (51)$$

The key dependences of the *NETD* emerge from Eq. (51):

- Sensor: The influence of the sensor is provided by the specific detectivity  $D^*$  and the electronic bandwidth  $B$ .
- Measurement object: The  $NETD$  is influenced by the temperature and the spectral specific radiation of the measurement object. The higher the temperature of the object measured is, the lower is the  $NETD$ .
- Optics: The f-number  $k$  of the optical system and the magnification, represented in Eq. (51) by the pixel grid  $A_P$  are also important factors in determining the  $NETD$ . The aperture  $k$  must be as small as possible, which is why wide-opening optics are primarily used for uncooled thermal imaging cameras.

In order to be able to interpret the  $NETD$  information of a thermal imaging camera, the temperature of the object measured and the f-number always have to be provided. In this regard, standard values are often assumed, for example, an object temperature of 27 °C and an f-number  $k = 1$  – irrespective of the optical system actually used!

The absolute measurement error specifies the permissible deviation from the true object temperature. A typical value for uncooled thermal imaging cameras is  $\pm 2 \text{ K} \pm 2\%$  of the reading.

The spatial resolution is described by optical-geometrical information and the  $MTF$ . Important optical-geometrical values are the opening angle ( $FOV$ : field of view), the resolution (number of pixels in the horizontal and vertical direction), the minimum geometrical resolution ( $IFOV$ : instantaneous field of view), and other information on the optics, such as the zoom range and factor, minimum focus distance, etc.

The magnitude  $L$  of the scenes can be calculated from the field of view, as follows:

$$L = 2R \tan \frac{FOV}{2}, \quad (52)$$

where  $R$  is the distance to the scene. The minimum geometrical resolution  $IFOV$  describes the magnitude  $l$  of a pixel in the scene:

$$l = 2R \tan \frac{IFOV}{2}. \quad (53)$$

The  $MTF$  of a thermal imaging cameras is essentially determined by the optical system and the geometrical  $MTF$  of the sensor. In cooled photon sensors, it is also possible for crosstalk to occur between pixels as a result of photon tunneling to the adjacent pixel.

The complicated technology required for manufacturing IR imaging sensors causes further problems, namely, “dead” pixels, a non-correctable characteristics inhomogeneity, and the long-term stability.

Dead pixels are non-functioning pixels, which are replaced by adjacent pixels. In particular, adjacent dead pixels (so-called clusters) can generate interference from

image artifacts. In thermal imaging sensors, up to 1% of dead pixels are generally permissible.

Each pixel possesses its own individual characteristic. Although all the pixels are corrected to a uniform characteristic using complicated algorithms, a small inhomogeneity nevertheless remains. This is described by the inhomogeneity-equivalent temperature difference *IETD*. It represents the extent of the correctability. The inhomogeneity in the image is also referred as the “fixed pattern.”

The long-term stability and service life of infrared sensors – and particularly coolers – have to be taken into account when using thermal imaging cameras.

---

## Passive Thermographic Techniques

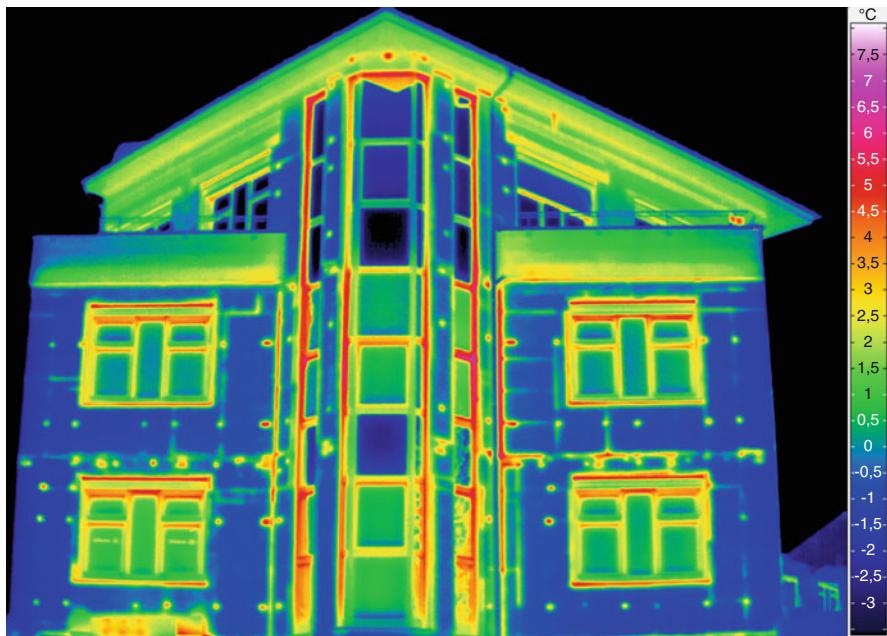
Passive thermographic techniques take advantage of the “natural” temperature differences present in a sample. These include temperature differences caused by the objects self-heating, e.g., the power dissipation of electronic components, or by friction.

When evaluating thermal images, it should be noted that the temperature differences measured for the object arise as a result of a local temperature distribution and differences in emissivity. For a different location-dependent emissivity, different radiances occur – even at a uniform, constant temperature, these being detected as apparent temperature differences. These are caused by reflections of the ambient radiation. The cause for the difference in temperature measured is not detectable in the thermal image and can only be determined in a specific evaluation of the image. For example, it is possible to eliminate the influence of the emissivity to a large extent with the help of a difference image. An example for the exploitation of emissivity differences is provided by the surface inspection. The emissivity of metal surfaces depends – amongst other things – on the roughness and oxidation state.

In addition to applications in nondestructive testing, thermography is primarily used in the construction industry (building thermography, Fig. 21) and in human and veterinary medicine (Vollmer and Möllmann 2010).

## Preventive Maintenance

Malfunctions and failures of machines and systems are often signalled at an early stage in the form of heat build-up and can be reliably detected by thermography. A particular benefit provided is that the measurement can usually be made without switching off or interrupting the operation of a system. The measurement ensues without any contact with the measuring point, thus enabling the procedure to be carried out from a safe distance or on objects in motion. Furthermore, a regular inspection of systems allows servicing and maintenance work to be performed in a planned manner. In order to detect defects, the comparison is made to a normal state,



**Fig. 21** Background-hidden structures are visible in a thermogram. (Courtesy of Infratec GmbH Dresden, Germany)

an asymmetric heat build-up or the exceeding of thresholds. Examples of preventive maintenance are:

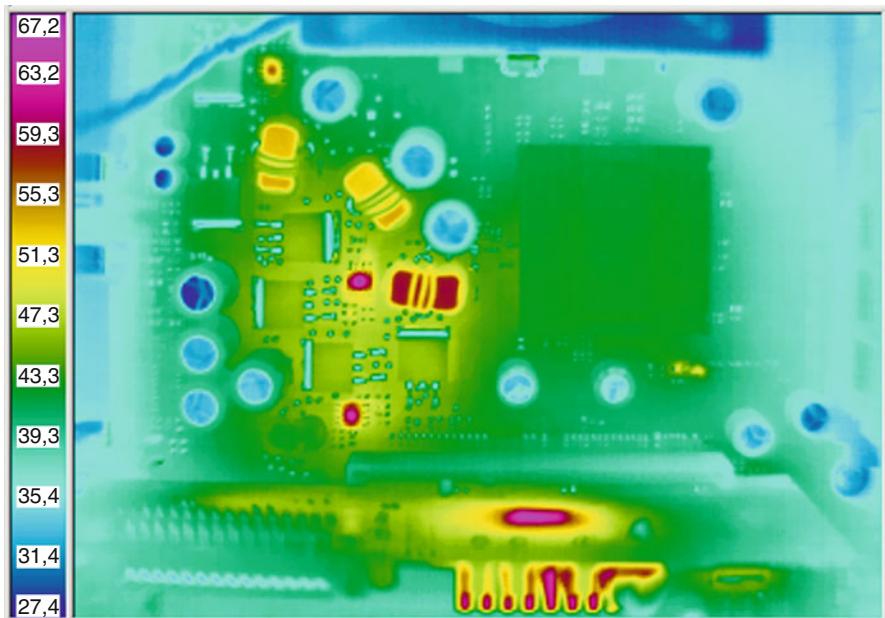
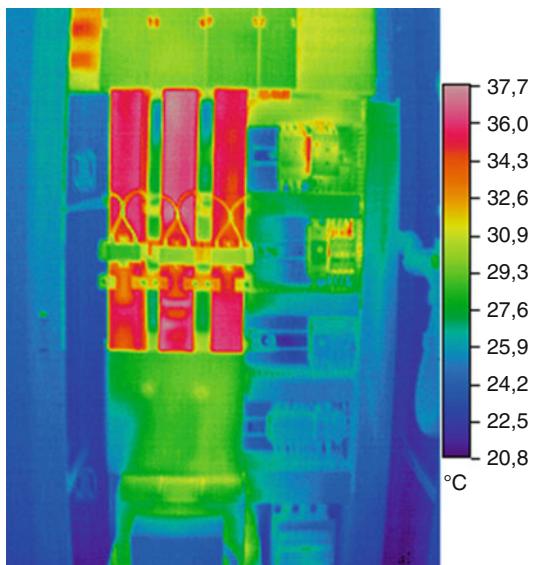
- All types of electrical systems, such as control cabinets, transformers, insulators, clamping connections, and power lines (Fig. 22)
- Cement rotary kilns
- Wheel bearings and tyres of road and rail vehicles
- Wear on mechanical assemblies, such as conveyor belts, turbines, etc

## Testing Electronic Components and Assemblies

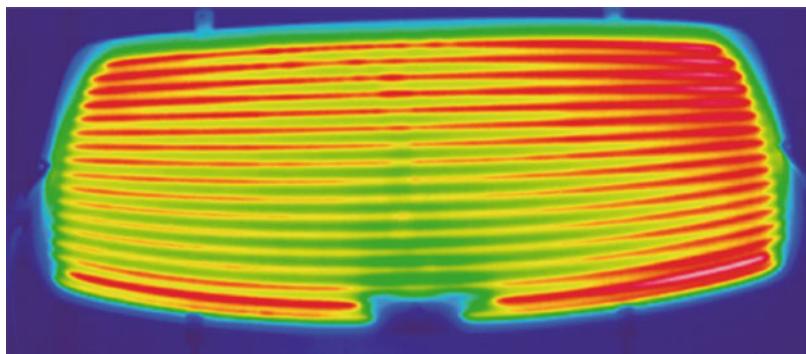
Thermography is used both in the development and testing of electronic assemblies. Thermal optimization, in particular, is a key problem in the development of new products. The thermal behavior of printed circuit boards can be well-investigated using thermographic systems (Fig. 23). Overheated electronic components, for example, are visible in the thermal image. The unfavorable arrangement of components that warm up can be readily identified, if they result in so-called hot spots, which, say, impair their reliability.

Macro lenses allow the resolution of extremely small structures (approx.  $>20 \mu\text{m}$ ). This also enables thermal measurements to be performed on microsystems.

**Fig. 22** Thermogram of hot contacts in a switch case.  
(Courtesy of DIAS Infrared  
GmbH Dresden, Germany)



**Fig. 23** Thermogram of printed circuit board. (Courtesy of DIAS Infrared GmbH Dresden,  
Germany)



**Fig. 24** Online temperature monitoring during the heat treatment of glass. (Courtesy of DIAS Infrared GmbH Dresden, Germany)

## Process Monitoring

The continuous measurement in production processes (so-called online thermography) is used for control and regulation purposes as well as for monitoring quality control in the manufacturing process (Fig. 24). Thermographic systems are used for non-contact measurements in automated temperature control systems. Examples are the monitoring and controlling

- Of chemical reactions
- Of drying processes
- In the production of plastics
- In metallurgy
- Automatic welding processes

The evaluation of online data is performed with industrial image processing methods and processes.

---

## References

- Budzier H, Gerlach G (2011) Thermal infrared sensors. Wiley, Chichester  
Rogalski A (1995) Infrared photon detectors. SPIE Press, Bellingham  
Vollmer M, Möllmann K-P (2010) Infrared thermal imaging. Weinheim, Wiley-VCH  
Williams TL (2009) Thermal imaging cameras. CRC Press, Boca Raton



# Active Thermography

38

Helmut Budzier and Gerald Gerlach

## Contents

Introduction .....	1402
Function and Design .....	1402
Thermographic Methods .....	1402
Energy Sources (Excitation Sources) .....	1406
Overview .....	1406
Energy Input with External Sources .....	1406
Pulse Thermography and Pulsed Phase Thermography .....	1408
Pulse Length and Penetration Depth .....	1408
Detection of Imperfections .....	1410
Evaluation of the Signal .....	1410
Lock-In Thermography .....	1412
Principle .....	1412
Heating the Test Specimen .....	1413
Lock-In Correlation Method .....	1413
Examples .....	1417
References .....	1418

## Abstract

In active thermography, heat is directed into a test object in order to detect inhomogeneities and cavities. When a test object is heated or cooled, temperature differences at the surface are caused due to local variations in the thermal conductivities and heat capacities of the test specimen. These are recorded and evaluated using a thermography system. The most important methods used in this process are pulse thermography, pulsed phase thermography, and lock-in thermography.

---

H. Budzier · G. Gerlach (✉)

Institute of Solid-State Electronics, Technische Universität Dresden, Dresden, Germany  
e-mail: [Helmut.Budzier@tu-dresden.de](mailto:Helmut.Budzier@tu-dresden.de); [Gerald.Gerlach@tu-dresden.de](mailto:Gerald.Gerlach@tu-dresden.de)

This paper first describes the basic measurement set-ups and the functional principles. Then possibilities of how to heat the test specimens are explained. External energy sources, such as radiation, convection, and heat conduction, heat the surface. With internal energy sources such as ultrasonic excitation, microwaves and inductive stimulation, in contrast, the heating takes place in the volume of the test specimen. If the specimen is excited with a short pulse, then this is referred to as pulse thermography. A special evaluation of the pulse response is pulsed phase thermography. If the excitation ensues in a sinusoidal form, then this is known as lock-in thermography.

---

## Introduction

### Function and Design

Active thermography is a thermographic process, in which a test specimen is taken out of its state of thermal equilibrium by the specific application of heat and the surface temperature is measured in the process. Due to local variations in the thermal conductivities and heat capacities of the specimen, temperature differences at the surface occur in the heating and cooling process. These are recorded and evaluated using a thermographic system (Fig. 1). The heat flow thus occurring in the specimen causes internal structures to become visible, if the individual structural elements possess different thermal characteristics (thermal conductivity, heat capacity) and dimensions. This enables inhomogeneities, delaminations, cavities, etc. to be detected.

The principal measurement set-up (Figs. 2 and 3) consists of:

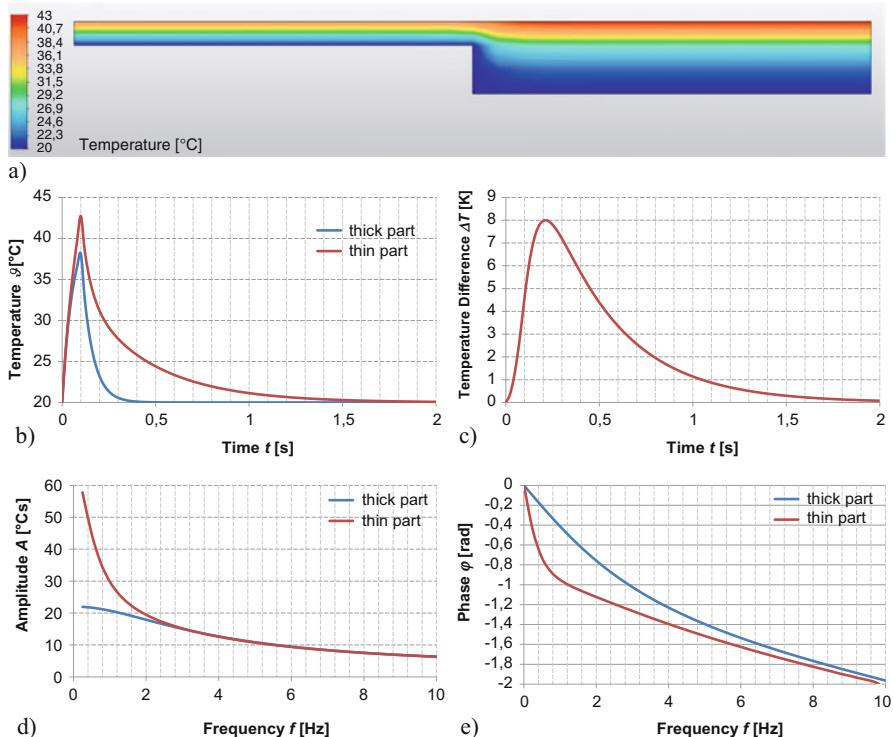
- An energy source, which heats the test specimen either periodically or with a short pulse
- A thermographic camera, which measures the surface temperature on a time-triggered basis
- A control unit to synchronize the measurements
- A computer for evaluating the data

### Thermographic Methods

In active thermography, two types of methods are distinguished: pulse thermography and lock-in thermography.

In pulse thermography, the test specimen is only subjected to a short energy pulse. The following evaluations ensue:

- The change in temperature as a function of time (Fig. 1b)
- The difference to a reference point or reference image (Fig. 1c)



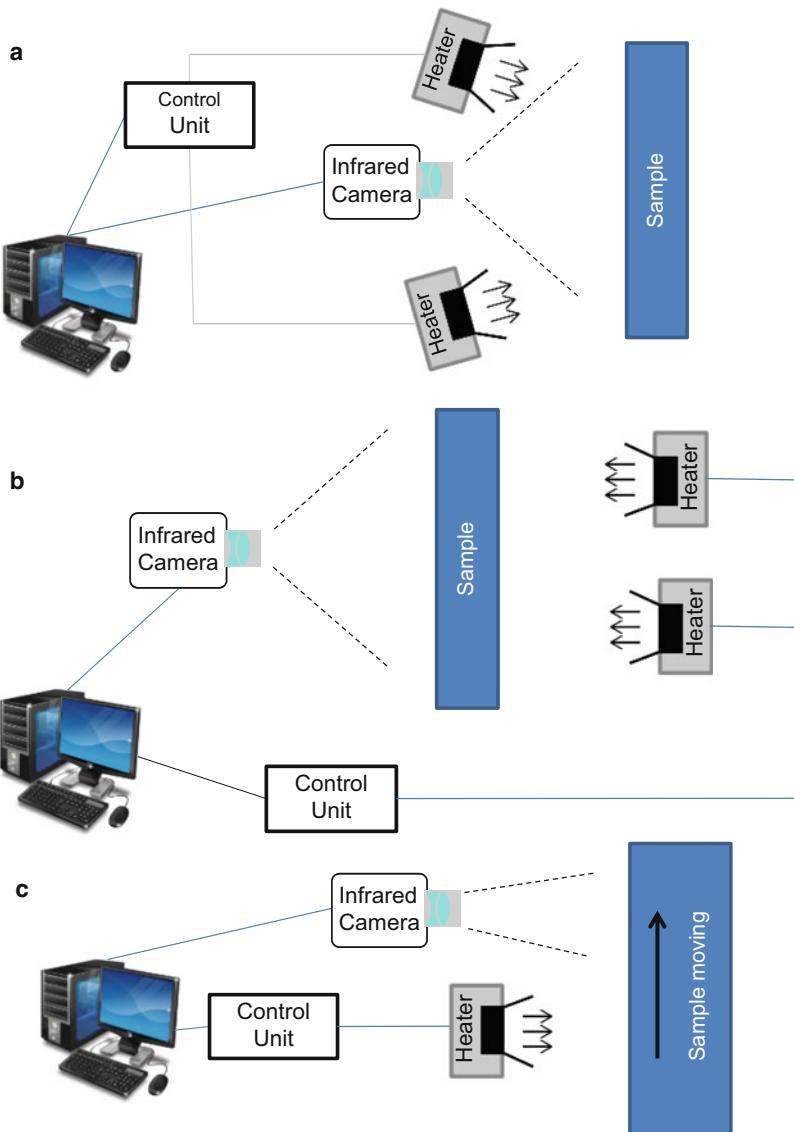
**Fig. 1** Pulse thermography: (a) the temperature distribution on a stepped aluminum body (3 or 9 mm thick) after 0.1 s (energy input at the surface of 0.0 s to 0.1 s); (b) the temperature gradient at the surface; (c) the temperature difference between the thick and thin part; (d) the amplitude response, and (e) phase response, calculated from the temperature gradients

- The change in temperature in the frequency domain (amplitude and phase response) (Fig. 1d, e).

The latter case is referred to as pulsed phase thermography. A special case of pulse thermography is when the pulse is of very long duration. This is called step thermography.

With lock-in thermography, the test specimen is subjected to the energy supplied by the lock-in frequency  $f_{LI}$  on a periodic basis, generally as a series of pulses. The progression over time of the change in temperature in the frequency domain (amplitude or phase) is then evaluated. In order to obtain, say, depth information on defects, the lock-in frequency can be varied. The lock-in thermography is then referred to as modulated thermography.

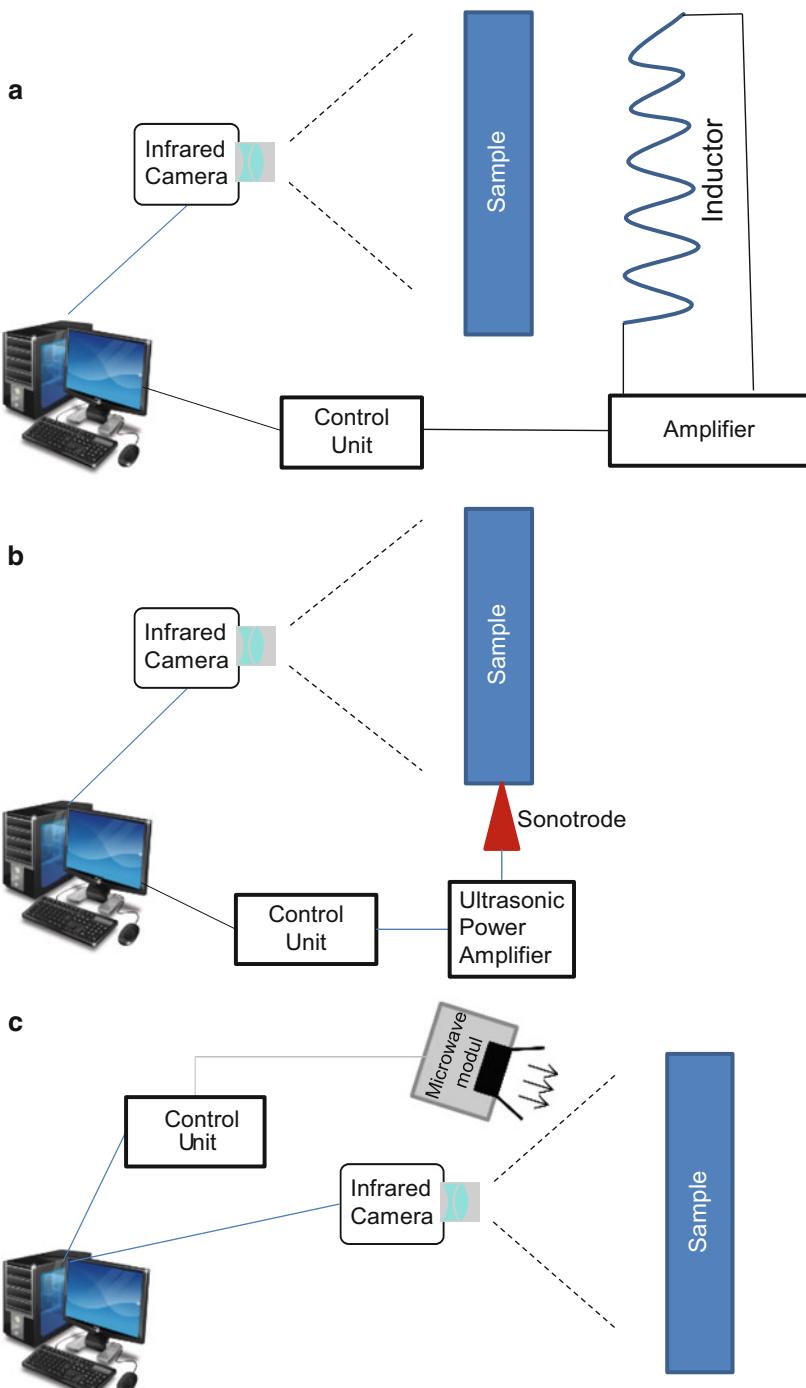
The results obtained by pulse and lock-in thermography are essentially the same. The advantage of pulse thermography is primarily to be found in the short measuring period for a comparably simple measurement set-up. Whereas lock-in thermography,



**Fig. 2** Measurement set-up with radiation excitation: (a) front-side thermography (in reflection), (b) flow thermography (in transmission), and (c) online variants of pulse thermography

in contrast, possesses a greater detection sensitivity in association with a significantly longer measuring duration.

If only defects are detected by the measurements, this is known as qualitative thermography. With quantitative thermography, in contrast, the defects are also characterized. It is possible to determine the size, position, and the thermal properties of defects.



**Fig. 3** Measurement set-up with internal excitation: (a) inductive excitation, (b) ultrasonic coupling, and (c) microwave excitation

The following paragraphs examine the energy input, pulse thermography, pulsed phase thermography, and lock-in thermography in detail and examples are presented.

## Energy Sources (Excitation Sources)

### Overview

The energy required to create a temperature gradient (or heat flow) can be provided in various ways (Table 1). The higher the energy provided, the greater are the temperature differences achievable in the test specimen. These, however, are limited by the maximum permissible temperature of the test specimen.

### Energy Input with External Sources

The classical methods of active thermography are based on an energy input from external sources (radiation, convection, and thermal conduction). In these processes, the surface of the test specimen is heated. This takes place either on the front side (reflectance measurement, Fig. 2a) or on the rear side (transmittance measurement, Fig. 2b) of the specimen. The latter is particularly favorable with thin specimens. If the front side is radiated, it should be noted that any reflections occurring may damage the thermographic camera. It may be necessary to dim the camera during the radiation process. Taking measurements during the heating period is therefore not usually possible. These methods are limited by a maximum permissible surface temperature of the test specimen and the inertia of the energy source (minimum pulse duration or maximum modulation frequency). It also has to be borne in mind that the

**Table 1** Heat-providing possibilities in test specimens

External sources	Radiation	Flash and photolamps Infrared emitters Halogen lamps Lasers Microwaves (electrically conductive materials) Sun (day/night change)
	Convection	Hot or cold air blowers Driers
	Thermal conduction	Heating plates Hot or cold water
Internal sources		Resistance heating (electrically) Inductive heating Sonic or ultrasonic excitation Chemical reaction Thermoelectric effect Microwaves (dielectrics)

test specimen must be able to absorb the energy. This means that it must possess a high absorption (or high heat transfer) coefficient.

When using internal sources, the heating ensues in the volume (body) of the test specimen. In general, inhomogeneities in the test specimen, such as cracks or imperfections (defect-selective method), are heated here. These are then readily discernible and can often be more easily detected than when using external sources.

The energy input can ensue in a variety of ways:

- Point form: Heating takes place on a local scale with heat flow both into the depth and to the side. This two-dimensional, symmetrical heat flow also enables cracks perpendicular to the surface to be detected. The entire test specimen is recorded by scanning.
- Linear: Heating is primarily used in online thermography. The measurement of the surface temperature ensues directly behind the line (pulse thermography).
- Planar: The entire surface of the test specimen is heated and thus recorded at the same time. This results in a one-dimensional heat flow into the depth (of the specimen). Uniform heating of the surface is extremely difficult to realize in practice, a fact that has to be taken into account in the evaluation.

### **Inductive Heating**

Inductive heating ensues from eddy currents in metals exposed to an alternating magnetic field (Fig. 3a). The Joule losses caused by the eddy currents heat the material internally. This thermal excitation by means of eddy currents possesses the advantage that it takes place very quickly without contact and can be automated. The energy input can be controlled by the geometric design (size and form) of the stimulator coil and thus adjusted to a given workpiece.

### **Heating by Ultrasound**

When heating using ultrasound (vibrothermography), advantage is taken of the elastic properties of the test specimen in order to conduct mechanical waves in the material (Fig. 3b). At places of imperfections, such as cracks and delaminations, friction – and thus heat build-up – occurs at the flanks of the defect, which can be measured on the surface. The geometry of the specimen can cause superimposition or elimination of the ultrasonic waves. This can be prevented by frequency modulation of the ultrasonic wave. Vibrothermography is a contact process. The coupling takes place using a sonotrode, attached to a specific place on the test specimen. It is important to take into account that at the coupling point, a high energy density occurs in the material of the test specimen, which must not damage the material. A further disadvantage with this method is the fact that it is scarcely possible to test complicated geometrical shapes using this method.

### **Heating Using Microwaves**

When using microwaves /Zhang/ to provide the energy input, surface or volume heating can occur – depending on the material (Fig. 3c). For conducting materials, such as metals and carbon-fiber composites, the electromagnetic wave does not

penetrate into the material but heats the surface by means of induced eddy currents. With dielectric materials, such as glass fiber composites, the body volume is heated by dielectric losses. The choice of suitable microwave transmitters, such as horn antennae, parabolic aerials, or waveguides, allows the heating of both small and very large objects, such as reinforced concrete walls.

### Heating Using Lasers

In laser thermography, the heating of the test specimen mostly ensues in a point-form manner by using lasers (or laser diodes) on the surface. The choice of a suitable beam guidance enables a component to be scanned in any manner desired (flying-spot laser thermography). A beam expansion, say, to the line, enables lasers to be used to great benefit in online thermography. Lasers allow the application of large amounts of power. Not only extremely short pulses but also very stable periodic excitations (in terms of time) can be created.

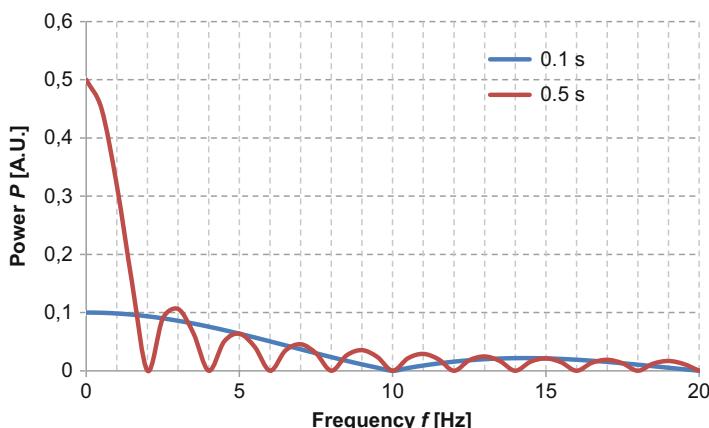
---

### Pulse Thermography and Pulsed Phase Thermography

In pulse thermography, the test specimen is excited by means of a short, high-energy pulse. The subsequent compensatory process, including the excitation process as appropriate, is recorded with a thermographic camera.

### Pulse Length and Penetration Depth

The length of the pulse has a decisive impact on the result of the measurement. The power spectrum of a pulse (Fig. 4) depicts the distribution of the energy supplied across the frequency spectrum. Short pulses have a wide distribution of



**Fig. 4** Power spectrum of a pulse; the parameter is the pulse length

energy, while long pulses provide significantly more energy at low frequencies. The thermal waves penetrating into the specimen body have the same frequency spectrum as the excitation pulse. The term “thermal wave” denotes both the time- and location-dependent periodic heating and cooling in the body. This process, however, is not based on the wave equation and is – from a strict physical perspective – therefore not a wave but a strongly damped cyclic process reaching into the depth of the body.

The thermal diffusion length  $\mu$  is the penetration depth of a thermal wave, for which the amplitude has declined to  $1/e$ :

$$\mu = \sqrt{\frac{\alpha}{\pi f}}. \quad (1)$$

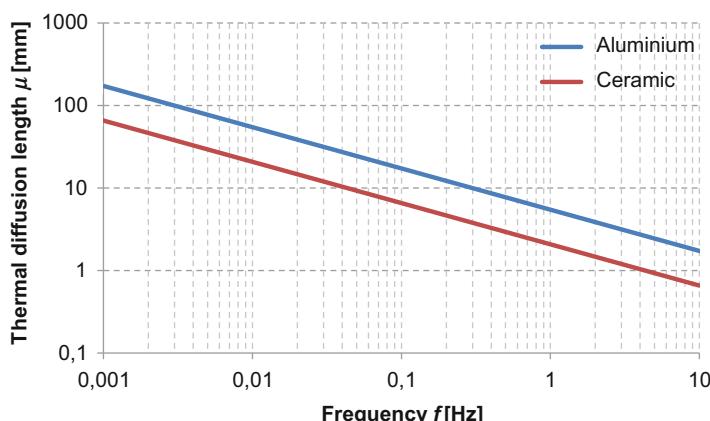
Here the thermal diffusivity  $\alpha$  is defined as:

$$\alpha = \frac{\lambda}{\rho c_p}, \quad (2)$$

where  $\lambda$  is the thermal conductivity,  $\rho$  the density, and  $c_p$  the specific heat capacity.

The thermal diffusion length therefore falls with increasing frequency. Figure 5 depicts the thermal diffusion length of a good (aluminum) and a poor thermal conducting material (ceramic). For the depth range of a measurement, the (1..2)-fold factor of the thermal diffusion length  $\mu$  is usually assumed.

If it is intended for the heat to penetrate deep into a body, low excitation frequencies (long pulse) and a greater measurement period are therefore required. For the measuring period  $t$ , the approximation applies as follows / Mal dague/:



**Fig. 5** Thermal diffusion length versus excitation frequency for different materials

$$t \cong \frac{z^2}{\alpha}, \quad (3)$$

where  $z$  is the depth. In the process, the contrast  $C$  strongly decreases with the depth  $z$  /Maldague/:

$$C \cong \frac{1}{z^3}. \quad (4)$$

## Detection of Imperfections

In order to detect imperfections, the thermal effusivity

$$b = \sqrt{\rho \lambda c_p} \quad (5)$$

of the imperfection must differ from that of the surrounding material. If the thermal effusivity of the imperfection is less than that of the surrounding material, then the imperfection will appear at the surface as an area of increased temperature.

## Evaluation of the Signal

To evaluate the pulse response, a series of thermal images is recorded. These thermal images show the progression of the surface temperature  $T(x, y, z = 0, t)$ . The data analysis ensues with the typical image processing methods. For this purpose, contrast-based procedures and a method using signal transformation are normally used.

### Thermal Contrast

The thermal contrast in the thermal image is not only created as a result of surface temperature differences but instead from differences in emissivity (see ► Chap. 37, “Passive Thermography, Thermal Imaging”). To this temperature differences are added arising from a nonuniform heating of the test specimen. This is why the direct evaluation of the contrast is often neither meaningful nor possible. Suitable methods are, for instance, difference imaging (comparison with nominal image or nominal point (Fig. 1c), running difference), background subtraction, and quotient formation.

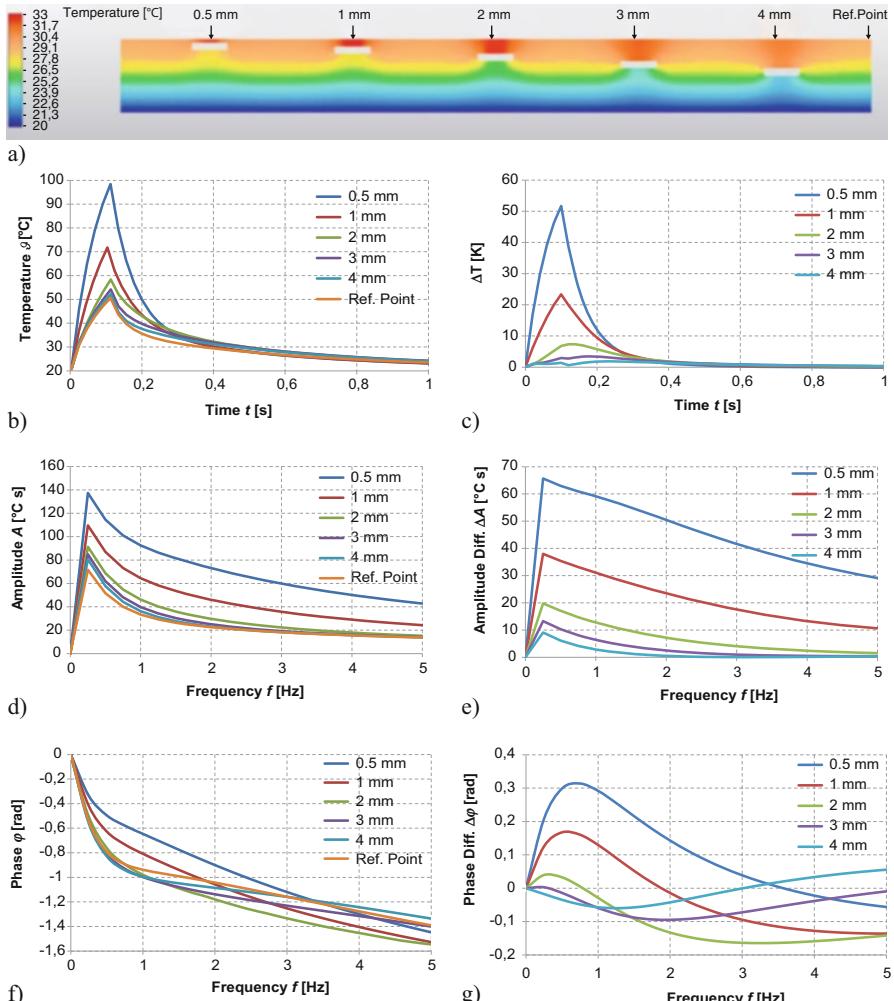
### Temperature Progression (Pulsed Phase Thermography)

Furthermore, it is possible to evaluate the temperature progression of an image point (pixel) over time. This applies to both the rise in the temperature curve and the time of maximum temperature. This evaluation can be optimally made using pulsed phase thermography.

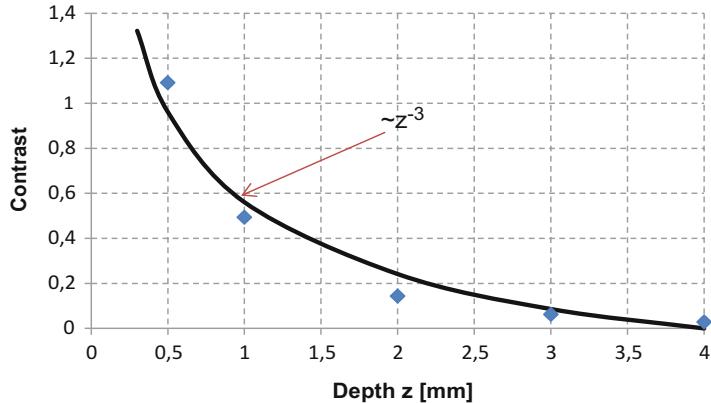
With pulsed phase thermography, a Fourier transform of the temperature progression over time is carried out for each pixel. The result is the amplitude and phase response (Fig. 1d, e). If the amplitude and phase responses of all pixels in an image are now combined, this is referred to as an amplitude or phase image.

Particularly favorable results are provided by the evaluation of the phase images. They reflect the progression over time of changes in temperature and basically represent a cross-section through the test specimen. Amplitude differences caused by, say, nonuniform heating of the specimen are not relevant. The depth range of the phase response is approx.  $2 \mu$ .

Figure 6 depicts an example of an aluminum body with cavities at various depths. It is clear that differential methods increase the detectability of cavities



**Fig. 6** Pulse thermography on an aluminum body with cavities at various depths: (a) the temperature distribution 0.2 s after an energy pulse of 0.1 s; (b) the temperature curve at the marked points above the cavities; (c) the temperature differences to a reference point; (d) the amplitude response; (e) the amplitude difference to the reference point; (f) the phase response; (g) the phase difference to the reference point; parameter is the depth  $z$



**Fig. 7** Contrast at the measuring points from Fig. 6b at the time of 0.1 s

(imperfections, inhomogeneities). Clearly evident too is how the contrast falls with increasing depth. Figure 7 shows the contrast in relation to the point of reference at the time of maximum contrast ( $t = 0.1$  s; at the end of the energy input). The empirical relationship from Eq. (4) has been included.

If the test specimen is briefly heated by an energy pulse, then it will no longer cool down to its starting temperature in the foreseeable future by adiabatic heating. This effect can significantly hamper the data evaluation, particularly the Fourier transform, and has to be taken into account – particularly with repeated measurements separated by short intervals.

If the idea of pulsed phase thermography is pursued and a test specimen is subjected to sinusoidal heat waves, we arrive at lock-in thermography.

## Lock-In Thermography

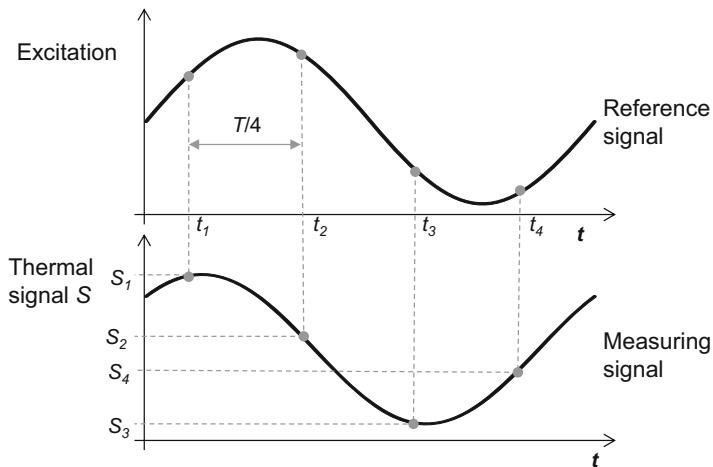
### Principle

In lock-in thermography, the test specimen is periodically heated with the frequency  $f_{LI}$ . A specific variation of the frequency  $f_{LI}$  results in a depth localization of inhomogeneities in the specimen.

If the object being measured is subjected to sinusoidal heating, then the amplitude and phase image can be calculated by four equidistant measuring points  $S_1 \dots S_4$  at the times  $t_1$ ,  $t_2 = t_1 + T/4$ ,  $t_3 = t_1 + T/2$ , and  $t_4 = t_1 + 3 T/4$  (Fig. 8). For the pixel  $n$  at the point  $(x, y)$ , the amplitude is given by:

$$A(n) = \sqrt{[S_1(n) - S_3(n)]^2 + [S_2(n) - S_4(n)]^2} \quad (6)$$

and the phase by:



**Fig. 8** Reference and measuring signal with the period  $T = 1/f_{Li}$  for a pixel

$$\varphi(n) = \arctan \left[ \frac{S_1(n) - S_3(n)}{S_2(n) - S_4(n)} \right]. \quad (7)$$

By taking the mean over numerous periods  $T$  of the measuring signal, it is possible to significantly improve the signal-to-noise ratio.

With the same test specimen as in the example of pulse thermography (Fig. 6), Fig. 9 shows the temperature curve of the lock-in thermography simulated using a lock-in frequency  $f_{Li}$  of 3 Hz

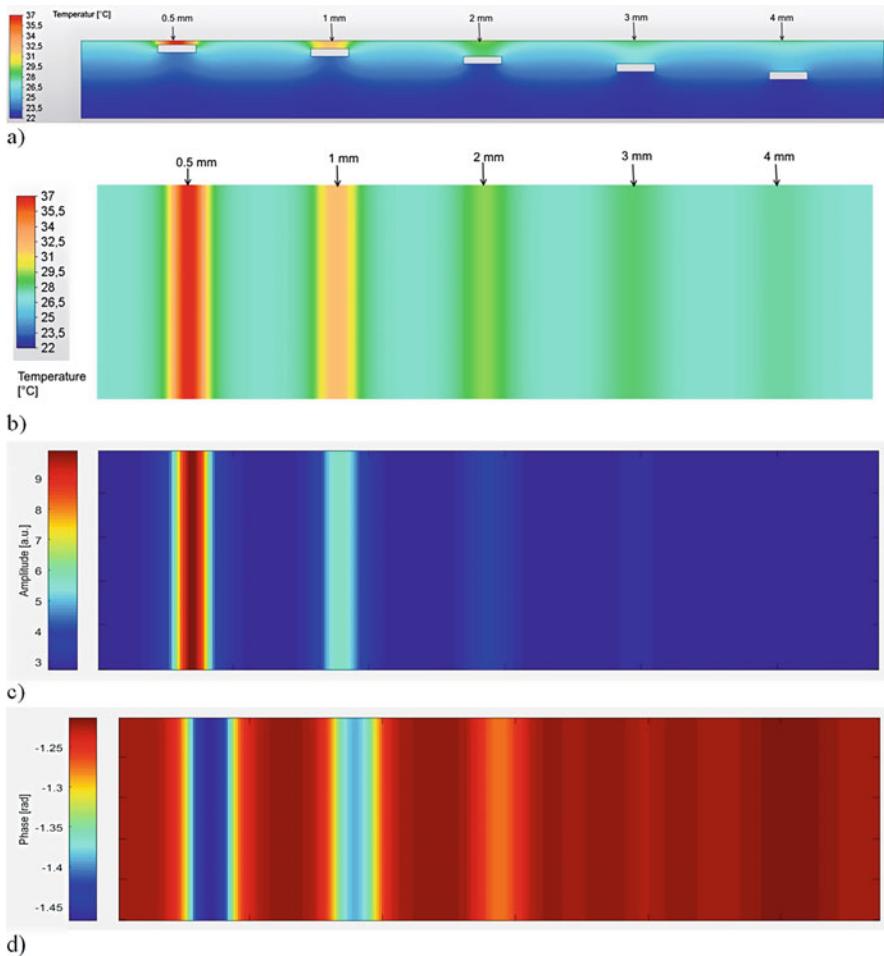
Figure 10 shows the amplitude and phase response at the surface. The penetration depth  $\mu$  is approx. 3 mm. The specific values of the phase are dependent on the starting time (here  $t_1$ , see Fig. 8). The depth of the imperfection can be determined by a variation of the lock-in frequency.

## Heating the Test Specimen

When calculating the amplitude and phase images, the adiabatic heating of the test specimen has to be taken into account. Figure 11 shows the simulated temperature progression. The test specimen is heated by the continuous energy input. Due to the large time constant, the increase in temperature can often be assumed as linear (blue curve in Fig. 11). This then allows a simple correction to the increase (Fig. 12).

## Lock-In Correlation Method

A general possibility of calculating amplitudes and phases is provided by the digital lock-in correlation. It can also be used if the heating of the specimen is not sinusoidal in form but ensues, say, by rectangular pulses. With the lock-in correlation, a periodically modulated measuring signal is weighted with a correlation signal of the same frequency,

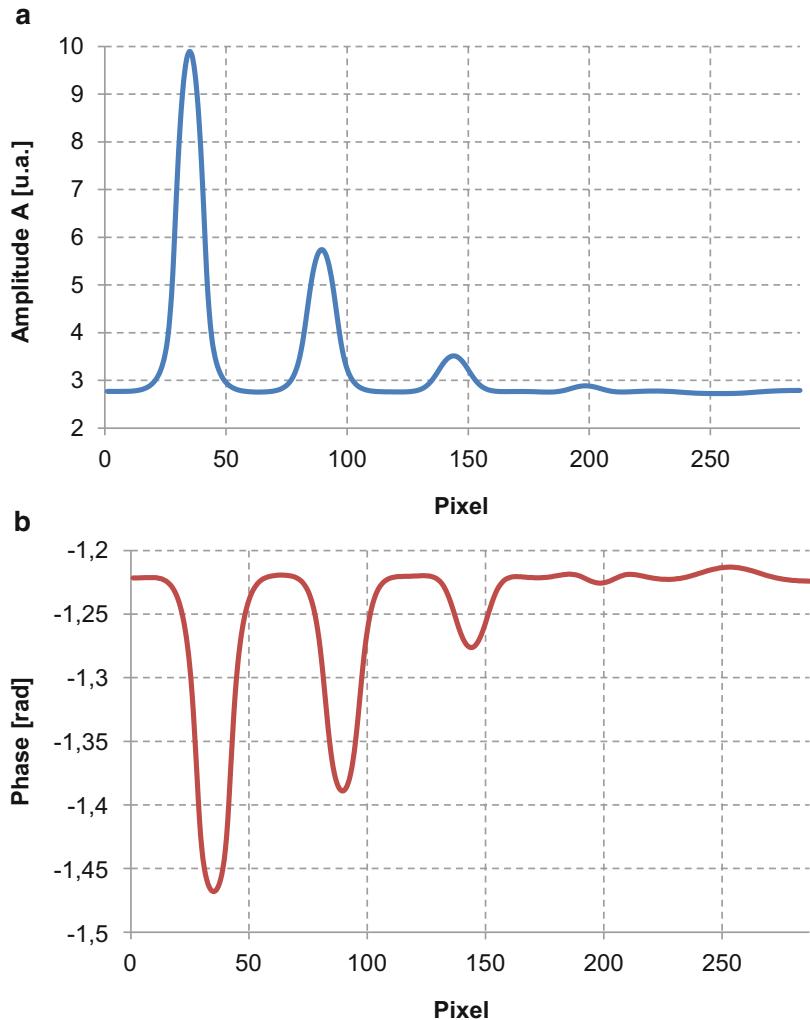


**Fig. 9** Lock-in thermography on an aluminum body with cavities at various depths after 0.55 s: (a) the temperature distribution at the cross-section; (b) the temperature distribution at the surface; (c) the amplitude image; (d) the phase image; lock-in frequency  $f_{LI} = 3$  Hz; parameter is the depth  $z$

e.g., of a sinus function. The mean value over time of the correlated signal is proportional to the amplitude of the carrier wave of the measuring signal. Interference signals, such as background noise, average out over the measuring period. For an unknown or variable phase, the correlation ensues with two orthogonal functions, say, a sine and cosine function. Figure 13 depicts the principle of the lock-in correlation.

For each pixel  $n$  at the point  $(x, y)$ , the weighted mean values  $S^{\sin}$  and  $S^{\cos}$  over a period  $T = 1/f$  of the excitation signal are first calculated from:

$$S^{\sin, \cos}(n) = \frac{1}{M} \sum_{m=0}^M K_m^{\sin, \cos} T_m(n), \quad (8)$$



**Fig. 10** (a) Amplitude response and (b) phase response at the surface of the aluminum body from Fig. 9

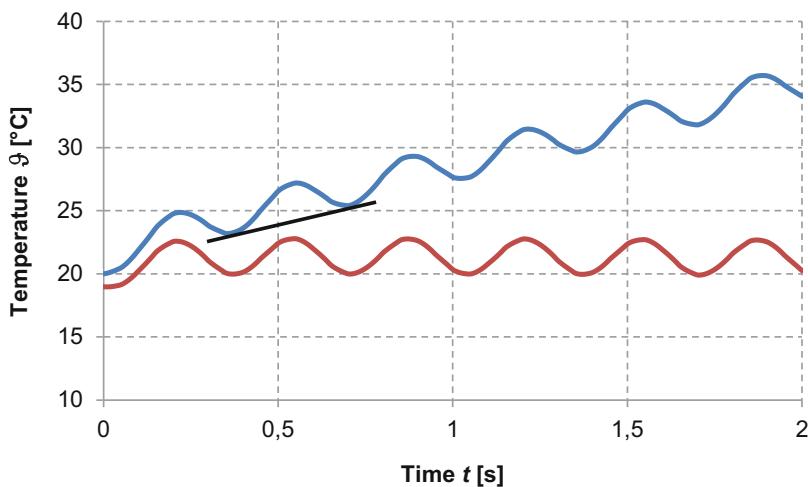
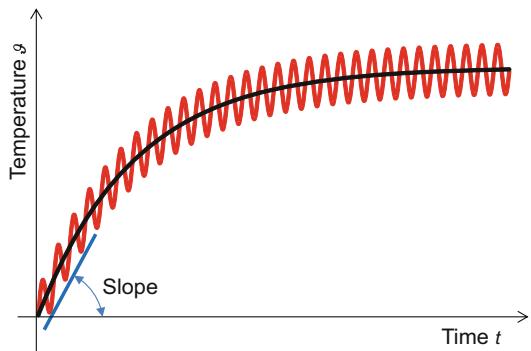
where  $M$  represents the measured values (thermal images) per period. For the correlation coefficients, the following applies:

$$K_m^{\sin} = 2 \sin \left( \frac{2\pi m}{M} \right) \quad (9)$$

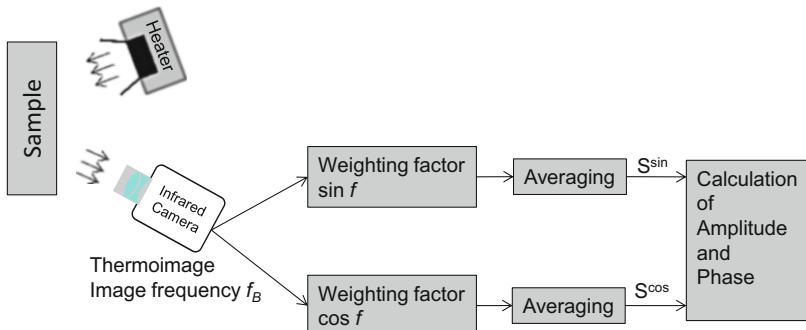
and

$$K_m^{\cos} = -2 \cos \left( \frac{2\pi m}{M} \right). \quad (10)$$

**Fig. 11** Principal temperature curve with lock-in thermography by heating the test specimen



**Fig. 12** Correction to the temperature increase of the test specimen from Fig. 11. Blue curve: measurement signal; black curve: correction line, calculated from the two minima touching the straight line; red curve: corrected temperature curve



**Fig. 13** Principle of the lock-in correlation method /Breitenstein/

The amplitude and the phase are then given by

$$A(n) = \sqrt{[S^{\sin}(n)]^2 + [S^{\cos}(n)]^2}, \quad (11)$$

$$\varphi(n) = \arctan\left(\frac{-S^{\cos}(n)}{S^{\sin}(n)}\right). \quad (12)$$

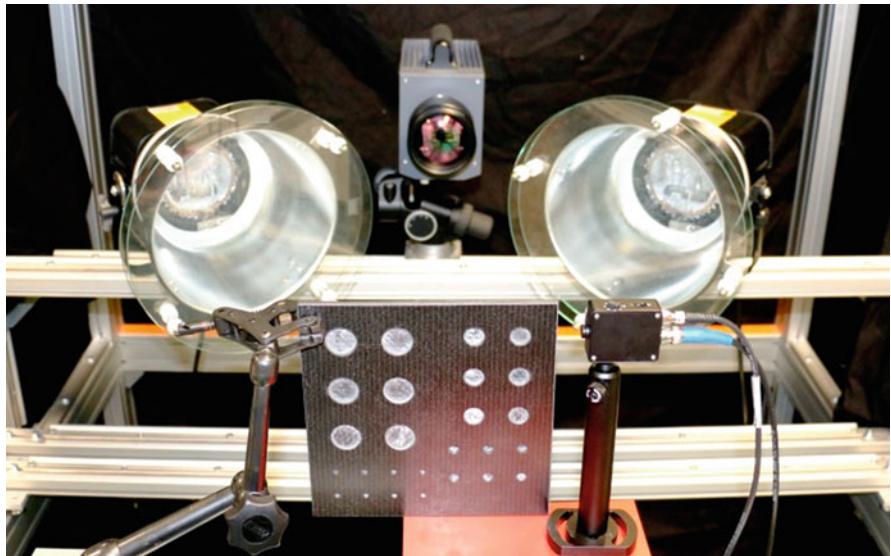
The number of measured values per period must be at least 4, requiring an image frequency  $f_B$  of

$$f_B \geq 4f. \quad (13)$$

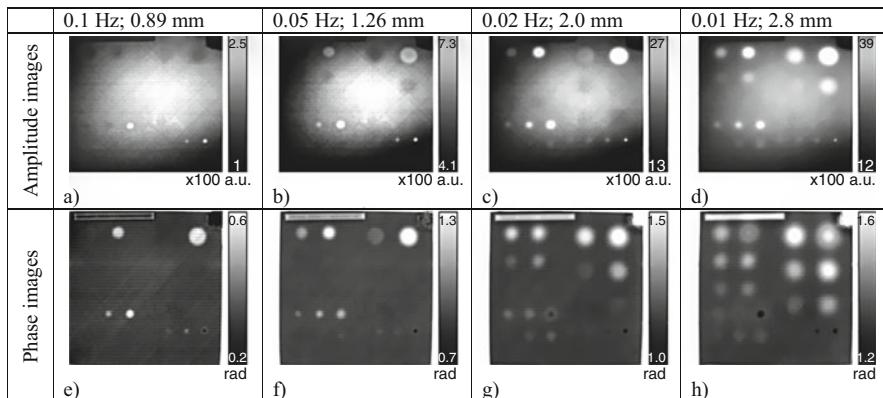
If the number of measurements is equal to 4, Eqs. (6) and (7) are then obtained. By averaging over many periods, it is also possible to significantly improve the signal-to-noise ratio here.

## Examples

Figure 14 depicts a test set-up for lock-in thermography /Maierhofer/. As the test specimen, a specially prepared CFRP (carbon fiber-reinforced plastic) plate can be seen in the foreground, in which blind holes of differing sizes and depths have been



**Fig. 14** Test set-up for lock-in thermography. The test specimen seen in the foreground is radiated by two halogen lamps with glass filters. In the background is a thermographic camera. (Courtesy of the Federal Institute for Materials Research and Testing (BAM) Berlin, Germany)



**Fig. 15** Amplitude images (**a–d**) and phase images (**e–h**) of the CFRP specimen from Fig. 14 at different lock-in frequencies. The figures above the images denote the lock-in frequency and the thermal penetration depth. (Courtesy of the Federal Institute for Materials Research and Testing (BAM) Berlin, Germany)

made from the rear. The glass filters in front of the halogen spotlights serve to suppress the infrared radiation of the spotlights, thus protecting the thermographic camera against reflections of infrared rays.

The result of a lock-in thermography using the test set-up from Fig. 14 can be seen in Fig. 15 /Maierhofer/. The nonuniform heating of the test specimen by the halogen spotlights is clearly detectable in the amplitude image. This is not the case in the phase image. In addition, phase images contain more detail. For this reason, the phase image is preferred for evaluating the measurements. The penetration depth (Eq. 1) varies as a function of the lock-in frequency. If the penetration depth is too low, inhomogeneities are not recorded. In comparison to Fig. 15e–h, this can be clearly recognized.

## References

- Breitenstein O, Warta W, Langenkamp M (2010) Lock-in thermography. Springer, Berlin  
 Maierhofer C, Myrach P, Krankenhagen R, Röllig M, Steinfurth H (2015) Detection and Characterization of Defects in Isotropic and Anisotropic Structures Using Lockin Thermography Journal of Imaging. <https://doi.org/10.3390/jimaging1010220>  
 Maldague X PV, Moore PO (eds) (2001) Nondestructive testing handbook. American Society for Nondestructive Testing, Columbus  
 Zhang H, Yang R, He Y, Foudazi A, Cheng L, Tian G (2017) A review of microwave thermography nondestructive testing and evaluation. Sensors 17(5):1123. <https://doi.org/10.3390/s17051123>



# Thermal Wave Techniques

39

Gunnar Suchaneck, Agnes Eydam, and Gerald Gerlach

## Contents

Introduction .....	1420
Basics of Thermal Waves .....	1420
Temperature Oscillations as Diffusion Waves .....	1420
Pyroelectric Coefficient .....	1422
Pyroelectric Current .....	1424
Pyroelectric Current Spectrum .....	1427
Analysis of Frequency Response .....	1428
Thermal Problem .....	1429
Radiation Loss, Relaxation Time, and Amplitude of Temperature Oscillation .....	1429
Boundary Conditions .....	1430
Multilayer Structures .....	1431
Lumped Model .....	1432
Thermal Pulse .....	1433
Thermal Steps .....	1435
Periodic Surface Temperature .....	1435
Thermal Methods .....	1436
Thermal Methods in the Frequency Domain .....	1437
Thermal Methods in the Time Domain .....	1444
Instrumentation .....	1448
Experimental Setup .....	1448
Examples of Nondestructive Evaluation of Embedded Piezoelectrics .....	1450
LTCC/PZT Sensor-Actuator Modules .....	1451
Macro-Fiber Composite Actuators .....	1455
Thermoplastic-Compatible Piezoceramic Modules (TPM) .....	1457
Integrated Cast Al/PI/PZT Piezomodules .....	1461
Examples of Thermal Pulse Measurements .....	1464

---

G. Suchaneck (✉) · A. Eydam · G. Gerlach

Institute of Solid State Electronics, Technische Universität Dresden, Dresden, Germany

e-mail: [gunnar.suchaneck@tu-dresden.de](mailto:gunnar.suchaneck@tu-dresden.de); [agnes.eydam@tu-dresden.de](mailto:agnes.eydam@tu-dresden.de);

[gerald.gerlach@tu-dresden.de](mailto:gerald.gerlach@tu-dresden.de)

---

Conclusion .....	1470
Summary .....	1471
References .....	1471

---

### Abstract

Thermal wave techniques can be used to excite thermal gradients within the volume to be evaluated and, hence, to achieve information on the interior of the sample under test.

The opportunities of thermal waves will be considered for the particular case of embedded piezoelectrics and the evaluation of their polarization state in construction elements, e.g., adaptronic structures. The described methods for this application are based on the pyroelectric effect whereas thermal excitation is induced by laser irradiation. Fundamentals of the pyroelectric effect are given and the resulting thermal problems are analyzed. Thermal methods recording the pyroelectric response both in the frequency (laser intensity modulation method including two- and three-dimensional polarization mapping) and in the time domains (thermal pulse method including thermal pulse tomography, thermal step method, thermal square wave method) are examined. Emphasis is given to the nonuniform resolution of thermal methods providing higher resolution in the near-surface region. Experimental pitfalls are highlighted. Nondestructive evaluation of piezoelectric transducers is illustrated by examples of lead zirconate titanate (PZT) plates embedded into low-temperature co-fired ceramics (LTCC), epoxy resins, thermoplastics, and die-casted aluminum.

---

## Introduction

Nondestructive examination (NDE) relies on the use of external excitations such as electromagnetic radiation, sound, temperature, etc., for obtaining local material properties without causing any damage to the investigated object itself. In this chapter, the charge generation by thermal fields is considered as a tool for NDE of materials, compounds, and electronic devices. Temperature oscillations penetrating into the sample give rise to local nonuniform thermal expansion causing mechanical strain  $x$ , local changes in dielectric permittivity  $\epsilon$ , or polarization  $P$ . This causes a pyroelectric current which carries information on the electric field or polarization profile. Thermal techniques are favorable for thin films. For relatively thick samples, polarization or electric fields should vary only with sample depth, otherwise the measurement becomes inaccurate. For data analysis, knowledge about thermal properties, e.g. the thermal diffusivity, is required.

---

## Basics of Thermal Waves

### Temperature Oscillations as Diffusion Waves

In the absence of internal heat sources, Fourier's law and the law of energy conservation result in a parabolic heat diffusion equation which yields the following for a homogeneous and isotropic material:

$$\Delta\Theta(z,t) - \frac{1}{D} \frac{\partial\Theta(z,t)}{\partial t} = 0, \quad (1)$$

where  $\Delta$  is the Laplace operator,  $\Theta(z,t) = T(z,t) - T_0(z,t_0)$  the temperature change of the temperature distribution relative to a reference temperature distribution, and  $D$  the thermal diffusivity. The solution of Eq. (1) for a slab with periodic surface temperature is a temperature oscillation which is attenuated and retarded in phase with increasing depth (Angström 1863). A harmonic temperature oscillation can be regarded as a highly damped wave exhibiting a complex wave vector determined by the angular frequency  $\omega$  of heat modulation (Carslaw and Jaeger 1959):

$$k = \sqrt{\frac{i\omega}{D}} = (1+i)\sqrt{\frac{\omega}{2D}}. \quad (2)$$

Note that following Fourier's law, the time average of the heat flux taken over a period is zero. Therefore, temperature oscillations are not traveling waves carrying energy, but just an oscillation of the temperature field (Salazar 2006). They neither possess wave fronts nor phase velocity and the square-law reflection-refraction principles must be replaced by field-gradient-driven accumulation-depletion rules (Mandelis et al. 2001). The main parameter of such a strongly damped wave is its penetration depth

$$d_D = \sqrt{\frac{2D}{\omega}}, \quad (3)$$

where the subscript  $D$  denotes diffusion since the oscillating solutions of Eq. (1) are part of a class of physical phenomena called diffusion waves (Salazar 2006; Mandelis et al. 2001). The penetration depths corresponding to maximum and minimum applied frequencies define the region of the sample from which information is gathered:

$$\sqrt{\frac{2D}{\omega_{\max}}} < z < \sqrt{\frac{2D}{\omega_{\min}}}. \quad (4)$$

Considering lead zirconate titanate (PZT) possessing a thermal diffusivity of  $0.5 \text{ mm}^2/\text{s}$  and a frequency range between 1 Hz and 1 MHz, the penetration depths of the temperature oscillation appear in the range of  $400 \mu\text{m}$  to  $400 \text{ nm}$ . Following Eq. (3), the temperature oscillation practically disappears within half a wavelength  $\lambda/2 = \pi \cdot d_D$ . The velocity of propagation of the temperature oscillations is given by (Göber et al. 1955):

$$v = \frac{\lambda}{T} = 2\sqrt{\frac{\pi D}{T}} = \sqrt{2\omega D}, \quad (5)$$

where  $T$  is the period of the temperature oscillation.

The portion of the sample thickness which is thermally activated by the modulated laser beam is approximately represented by a scanning function (Ploss et al. 1992; Lang 1998, 2004):

$$\varphi(z) \propto \sin\left(\frac{z}{d_D}\right) \exp\left(-\frac{z}{d_D}\right), \quad (6)$$

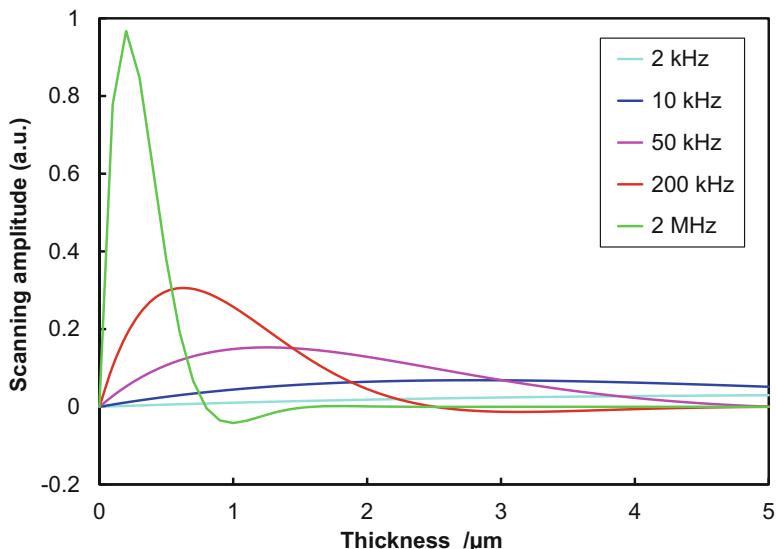
shown for PZT in Fig. 1.

Consequently, this can be used as a tool that provides an electrical response from the whole volume (at low frequencies) and from a surface layer (at high frequencies). This way, information is acquired from different parts of the sample.

## Pyroelectric Coefficient

The pyroelectric coefficient  $p$  at constant stress  $X$  and electric field  $E$  is the sum of the pyroelectric coefficient at constant strain  $x$  (primary effect) caused by the actual change of spontaneous polarization during temperature variations and the piezoelectric effect due to thermal strain (secondary effect) (Nye 1995):

$$p = \left(\frac{\partial P}{\partial T}\right)_{E,X} = \left(\frac{\partial P}{\partial T}\right)_{E,x} + \left(\frac{\partial P}{\partial X}\right)_{E,T} \left(\frac{\partial X}{\partial x}\right)_{E,T} \left(\frac{\partial x}{\partial T}\right)_{E,X}. \quad (7)$$



**Fig. 1** Spatial variation of the scanning function, Eq. (6), for PZT at different laser modulation frequencies

In anisotropic materials, pyroelectricity is a vector property. The vector components of the pyroelectric coefficient at constant electric field are then obtained by the following equation:

$$p_i^X = p_i^x + d_{ijk}^T c_{jklm}^T \alpha_{lm}^X, \quad (8)$$

where  $i = 1, 2, 3$  corresponds to the  $x, y$ , and  $z$  directions in the orthogonal Cartesian coordinate system;  $p_i^X$  and  $p_i^x$  denote the pyroelectric coefficient measured at constant stress and strain, respectively;  $d_{ijk}^T$  represents the third-rank piezoelectric tensor at constant  $T$ ;  $c_{jklm}^T$  is the fourth-rank elastic stiffness tensor at constant  $T$ ; and  $\alpha_{lm}^x$  is the second-rank thermal expansion tensor at constant stress. The primary pyroelectric effect – at constant volume and fixed shape – the first term on the right side of Eq. (8) of the material considered in this chapter,  $\text{PbZr}_x\text{Ti}_{1-x}\text{O}_3$  (PZT), is opposite in sign to the much smaller secondary effect where additional polarization changes occur in free-standing samples (second term on the right side of Eq. (8)) (Cook et al. 1963). In relaxor ferroelectrics like  $(1-x)\text{PbMg}_{1/3}\text{Nb}_{2/3}\text{O}_3 - x\text{PbTiO}_3$  (PMN-100xPT), the piezoelectric effect due to thermal strain provides a significant secondary contribution to the pyroelectric response (Movchikova et al. 2010). Also, secondary pyroelectricity can account for about half the pyroelectricity observed in polyvinylidene fluoride (PVDF) (Kepler and Anderson 1978).

An additional tertiary contribution to the pyroelectric effect arises in the presence of spatial strain gradients. For instance, inhomogeneous heating causes thermo-elastic stress which in turn induces polarization through the piezoelectric effect (Cady 1964; Schein et al. 1978; Nye 1995). Under the usual measurement conditions, when the electrodes completely cover the plane-parallel surfaces of the pyroelectric element and no mechanical load is applied to any of the surfaces, a residual self-balanced stress field remains (Braccini and Dupeux 2012). Here, the tertiary pyroelectric effect caused by local thermal stress is not observed since the contribution of the averaged thermal stress value is measured. If only a part of the surface is heated, a tertiary pyroelectric effect appears due to the presence of a temperature gradient at the edge of the heated region (Wang Xu-Sheng 1993). It is opposed to the primary and secondary ones and lowers the measured pyroelectric coefficient.

Symmetric tensors are usually represented by reducing their order (Voigt 1910). In this case, the double-subscript tensor notation of the thermal expansion tensor reduces to a single-subscript matrix notation, and both the triple-subscript tensor notation of the piezoelectric tensor and the quadruple-subscript notation of the elastic stiffness tensor simplify to a double-subscript matrix notation.

In the case of a thin film laterally perfectly clamped in the substrate plane, the pyroelectric coefficient is given by Lines and Glass (1977) and Zook and Liu (1978):

$$p_3 = p_3^X - \frac{2d_{31}\alpha_1}{s_{11} + s_{12}}, \quad (9)$$

where

$$p_3^X = p_3^x + \sum_j e_{3j} \alpha_j. \quad (10)$$

Here,  $s_{ij}$  are the elastic compliance constants (the inverse tensor of the elastic stiffness) and  $e_{ij}$  the piezoelectric stress constants related to the piezoelectric strain constants  $d_{ij}$  by  $d_{nj} = e_{ni}s_{ij}$ . Pyroelectric coefficients  $p_3 = p(z)$  for ferroelectric thin films on thin membranes determined by temperature oscillation methods lie within the limits between the constant strain and the ideal clamping cases.

The values of pyroelectric coefficients cited in literature for crystals and ceramics are usually for stress-free conditions. On the other hand, for ferroelectric pyroelectric thin films on a substrate, the effect of clamping on  $p$  is rather small.

In the presence of piezoelectric resonances, the pyroelectric coefficient is given by (Neugschwandtner et al. 2001):

$$p_3 = \frac{p_3^X + \frac{\beta_3^D e_{33}}{c_{33}^D} \frac{\tan\left(\frac{\pi f}{2f_{ta}}\right)}{\frac{\pi f}{2f_{ta}}} + \frac{\alpha_1^E d_{31}}{s_{11}} \frac{\tan\left(\frac{\pi f}{2f_l}\right)}{\frac{\pi f}{2f_l}} + \frac{\alpha_2^E d_{32}}{s_{22}} \frac{\tan\left(\frac{\pi f}{2f_w}\right)}{\frac{\pi f}{2f_w}}}{1 - k_{33}^2 \frac{\tan\left(\frac{\pi f}{2f_{ta}}\right)}{\frac{\pi f}{2f_{ta}}}}, \quad (11)$$

where  $\beta_j = \alpha_i/s_{ij}$  denotes the thermal stress coefficient,  $k_c$  the electromechanical coupling coefficients,  $f_{ta}$  the transverse anti-resonance frequency,  $f_l$  the length extension resonance frequency, and  $f_w$  the width extension resonance frequency.

## Pyroelectric Current

A pyroelectric material comprising two electrodes represents an electrical source. It is defined by its open-circuit voltage or its short-circuit current. Modulation of a constant intensity light source leads to an energy per modulation period that decreases with frequency. The resulting  $1/f$  frequency dependence of the recorded pyroelectric voltage (charge) signal may be eliminated by a time derivative, i.e., it is most advantageous to record the current response. Since the response time of the polarization is much smaller than the thermal relaxation time (of the order of picoseconds for polarization response (Perry et al. 1964; Liu et al. 2013), and nanoseconds for polar nanocluster flipping and breathing in relaxor ferroelectrics at room temperature (Petzelt et al. 2015)), the current produced by the primary pyroelectric effect will be proportional to the rate of temperature change. Assuming that the characteristic dimension of the thermally excited top electrode is much larger than the sample thickness, the transient temperature distribution across the sample thickness is a one-dimensional problem. Under short-circuit conditions and in

in the absence of an internal electric field, the complex pyroelectric current  $I_p(t)$  is described then by a fundamental relation (Lang and Das-Gupta 1986; Ploss et al. 1992):

$$I_p(t) = \frac{A}{d} \int_0^d r(z) \frac{\partial}{\partial t} \Theta(z,t) dz, \quad (12)$$

where  $t$  is the time,  $A$  the heated sample area,  $d$  the thickness of the piezoelectric sample,  $z$  the depth coordinate, and  $r(z)$  the pyroelectric response function,

$$r(z) = \alpha_P P(z) - (\alpha_z - \alpha_e) \varepsilon \varepsilon_0 E_{\text{int}}(z) = p(z) - (\alpha_z - \alpha_e) \cdot \varepsilon \varepsilon_0 E_{\text{int}}(z). \quad (13)$$

The latter is determined both by the pyroelectric coefficient  $p(z)$  and the internal electric field  $E_{\text{int}}$ . It is generated in a short-circuited sample ( $E = 0$  at electrode surfaces) due to space charges and the depolarization field

$$\begin{aligned} E_{\text{int}}(z) &= \frac{1}{\varepsilon_r \varepsilon_0} \\ &\cdot \left[ \left( \int_0^z \rho(\zeta) d\zeta - \frac{1}{d} \int_0^l \left( \int_0^{z'} \rho(\zeta) d\zeta \right) dz' \right) - \left( P(z) - \frac{1}{d} \int_0^l P(z') dz' \right) \right] \\ &= \delta E_{sc} - \frac{\delta P}{\varepsilon_r \varepsilon_0}, \end{aligned} \quad (14)$$

with  $\alpha_P$  the temperature coefficient of spontaneous polarization,  $\alpha_z$  the thermal expansion coefficient,  $\alpha_e$  the temperature coefficient of the dielectric constant,  $\varepsilon_r$  the relative dielectric permittivity,  $P(z)$  the spatially dependent spontaneous polarization, and  $\rho(z)$  the spatially dependent space charge density. For mathematical convenience, the limits of the integral in Eq. (12) may be extended toward infinity since  $r(z)$  has a non-zero value only inside the sample. Without a priori knowledge of the space charge distribution, a separation of the pyroelectric response and the space charge distribution is not possible by using the laser intensity modulation method (LIMM) alone (Lang 1990), i.e., the response function  $r(z)$  is not able to distinguish between contributions from space charge and polarization. Therefore, the internal electric field  $E_{\text{int}}(z)$  must be determined separately. Usually, the pyroelectric spectrum is analyzed assuming a local compensation of the polarization gradient by an appropriate charge density  $\rho = dP/dz$  (Ploss et al. 1992; Sandner et al. 2002). According to Eq. (14), this leads to the simple case  $E_{\text{int}} = 0$  and, consequently,  $r(z) = p(z)$ . Assuming a homogeneous polarization,  $p(z) = p_0 = \text{const.}$ , Eq. (12) simplifies to:

$$I(t) = \frac{A}{d} p_0 \int_0^d \frac{\partial}{\partial t} \Theta(z,t) dz = A p_0 \frac{d \langle \Theta \rangle}{dt}. \quad (15)$$

In the absence of primary pyroelectricity, but in the presence of an internal field, Eq. (12) reads:

$$I(t) = \frac{A}{d} (\alpha_e - \alpha_z) \varepsilon \varepsilon_0 \int_0^d E_{int}(z) \frac{\partial}{\partial t} \Theta(z,t) dz. \quad (16)$$

Equation (16) is used to analyze the electric field distribution  $E(x)$  and, thus, via Poisson's equation, the space charge accumulation phenomenon in dielectrics, e.g., in direct current cable insulations at high voltages, which cause their degradation. For space charge measurements, an overview of thermal methods can be found in Imburgia et al. (2016). Although this topic is beyond the scope of this chapter, issues of the thermal problem, data analysis, and spatial resolution could be of interest.

Unfortunately, Eq. (12) is a convoluted function (Fredholm integral equation of the first kind) of the temperature (Lang 2004). Therefore, cumbersome mathematical deconvolution techniques are required to compute the polarization, electric field, or space charge distributions from the time dependence of the pyroelectric current and the rate of temperature change and to interpret the results correctly. Here,  $\partial\Theta(z,t)/\partial t$  is usually calculated from the heat diffusion equation, Eq. (1).

Since the volume element at a distance of about the penetration depth  $d_D = (2Dt)^{1/2}$  provides most of the pyroelectric signal at a corresponding time  $t$  (Coufal et al. 1987), the pyroelectric response may be probed by pulse excitation using  $(2Dt)^{1/2}$  as a depth scale. This scaling transformation approach was first reported for electric field evaluation in Leal Ferreira (1989). Scaling transformation is also carried out in the frequency domain. According to Eq. (3), the sample depth region that is thermally probed becomes small at high frequencies. On the other hand, the solution of Eq. (12) in the frequency domain predicts a large pyroelectric response even in the vicinity of the penetration depth of the temperature oscillation. Here, the slope of the absolute value of  $\partial\Theta/\partial t$  has its maximum value (Lang 2004). Therefore, for both-side electroded piezoelectric ceramics probed by temperature oscillations in the surface region with a penetration depth much smaller than the sample thickness, an approximate profile of the pyroelectric coefficient is given by Ploss et al. (1992):

$$p(d_D) = \frac{cd}{\Phi_1 A} [\Re(I_p(\omega)) - \Im(I_p(\omega))], \quad (17)$$

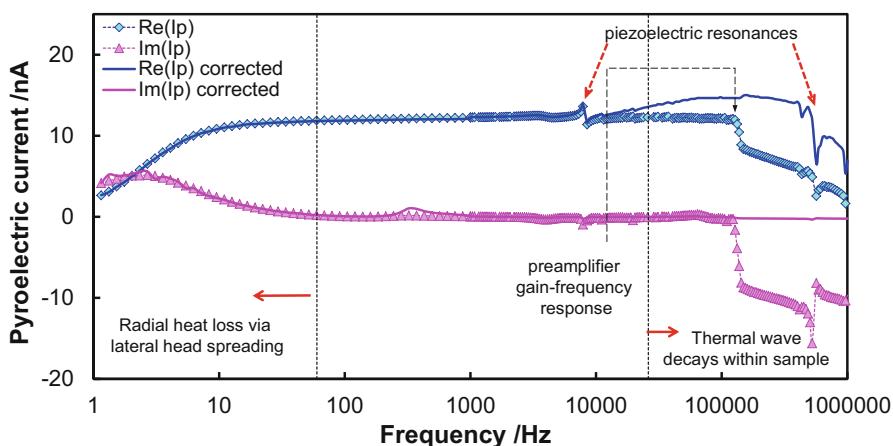
where  $\Phi_1$  is the heat flux absorbed at the top surface of the electrode,  $c$  is the volumetric specific heat, and  $\Re(I_p(\omega))$  and  $\Im(I_p(\omega))$  are the real and imaginary parts of the pyroelectric current, respectively. Equation (17) means that, at a given angular frequency  $\omega$ , the difference between the real and imaginary parts of the measured pyroelectric current transforms directly into the pyroelectric response function  $r(z)$ , at the penetration depth of a thermal oscillation, Eq. (3). It provides a sufficient approximation for about  $1/4$  of the material thickness near the upper surface. Here, the uncertainty of this approximation is less than 1%. In the middle of the sample, the

discrepancy with the exact solution increases up to 25% (Bauer and Bauer-Gogonea 2003). In practice, the upper frequency is limited by corrections due to the frequency-dependent gain and the phase shift of the preamplifier.

## Pyroelectric Current Spectrum

Figure 2 shows the pyroelectric current spectrum of a  $4\text{ }\mu\text{m}$  thick  $\text{LiTaO}_3$  pyroelectric element with an active area of  $2 \times 2\text{ mm}^2$  (Gerlach et al. 2007). At low modulation frequencies (below 60 Hz), heat loss caused by transverse heat conduction and heat loss to the surroundings should be taken into account. Medium frequencies ( $60\text{ Hz} < f < 25\text{ kHz}$ ) could be used to calibrate the LIMM signal to the averaged pyroelectric coefficient measured separately by other techniques. In this frequency region, the penetration depth of the temperature oscillation exceeds the sample thickness. On the other hand, the solution for an infinite solid starts to be valid above 50 kHz. At higher frequencies, the gain/phase characteristics of the preamplifier could become a crucial factor. Here, the measured LIMM spectra must be corrected by performing calibration with a fast photodiode.

The traveling temperature oscillation causes the sample to develop thermal stresses and strains. The strain produced by thermal expansion of the surface layer gives rise to acoustic waves. The sample is then forced into resonance vibrations (Von Laue 1925). As a consequence, piezoelectric bending oscillations appear superimposed on the pyroelectric signal. The first piezoelectric resonance at around 8 kHz is a bending oscillation visible also in measurements of the pyroelectric element acceleration sensitivity (Gerlach et al. 2004).



**Fig. 2** Pyroelectric current spectrum of a  $\text{LiTaO}_3$  pyroelectric element ( $d = 4\text{ }\mu\text{m}$ ,  $A = 4\text{ mm}^2$ ). Shown are distortions of the LIMM spectrum by the gain/phase characteristics of the preamplifier and piezoelectric resonances.  $\text{Re}$  and  $\text{Im}$  denote real and imaginary parts of the pyroelectric current, respectively (Gerlach et al. 2007). (Reprinted with permission of SPIE)

## Analysis of Frequency Response

Frequency response is the quantitative measure of the output spectrum of a system in response to a stimulus. It is used to characterize the dynamics of the system. In electrical engineering and control theory, Bode, Nyquist, and Nichols plots are graphs of the frequency response (D’Azzo and Houpis 1999). In our case, the parametric Nyquist plot is the most appropriate. Nyquist plots provide a graphic insight into the system under study and thus, a more direct and easy understanding of the heat conduction processes. On the other hand, a mathematical fit of the Nyquist plot to a model frequency response facilitates elimination of measurement artifacts arising inevitably in the picoampere range. For instance, the measurement uncertainty of the Solartron 1260 impedance/gain phase analyzer is 0.2 nA.

The frequency response of non-resonant, linear systems with a periodic pulse signal (signal amplitude  $F$ , zero-point equilibrium) is represented in the frequency domain by a function  $F(i\omega)$  at each modulation frequency  $\omega$  in terms of gain

$$|F(i\omega)| = \sqrt{\Re[F(i\omega)]^2 + \Im[F(i\omega)]^2} = \frac{F_{\text{out}}}{F_{\text{in}}}, \quad (18)$$

and phase shift

$$\phi = \arctan \frac{\Im[F(i\omega)]}{\Re[F(i\omega)]}. \quad (19)$$

Contrary to well-known dielectric relaxation (Kremer and Schoenhals 2003), a frequency response  $F(i\omega) = 1 - F(i\omega)_{\text{diel}}$  has to be considered since thermal losses decrease with increasing frequency whereas the pyroelectric response increases. A Debye relaxation function  $f_D(t) = \exp(-t/\tau)$ , where  $\tau$  is the single relaxation time, leads in the frequency domain to

$$F(i\omega)_D = \frac{1}{\tau} L \left[ \exp \left( -\frac{t}{\tau} \right) \right] = \frac{1}{1 + i\omega\tau}, \quad (20)$$

where  $L[f(t)]$  denotes the Laplace transformation. Correspondingly, one obtains

$$F(i\omega) = 1 - F(i\omega)_D = \frac{i\omega\tau}{1 + i\omega\tau}. \quad (21)$$

In the Nyquist complex plane mapping, Eq. (21) yields a semicircle where the frequency increases leading to higher real part values from the left to the right. More advanced relaxation models of, e.g., dielectric relaxation (Kremer and Schoenhals 2003), consider the superposition of individual relaxation processes which lead to a distribution of relaxation time constants. However, the integral of a superposition of Debye relaxations possesses solely approximate analytical solutions (Davidson and Cole 1951).

The empirical Havriliak-Negami equation derived by means of complex analysis (Havriliak and Negami 1967) accounts for the broadness and asymmetry of the dielectric dispersion curve by adding two empirical parameters  $\alpha$  and  $\beta$ . In our case, it is given by

$$F(i\omega) = 1 - \frac{1}{[1 + (i\omega)^a]^b}, \quad (22)$$

where  $a$  and  $b$  describe the broadness and the asymmetry of the spectrum, respectively. For  $b = 1$ , the Havriliak-Negami relaxation reduces to the special case of the Cole-Cole relaxation (Cole and Cole 1941). Another special case of the Havriliak-Negami equation ( $b < 1, a = 1$ ) is the Cole-Davidson relaxation (Davidson and Cole 1951). In this case, the parameter  $b < 1$  leads to a skewed arc of the semi-circle of the frequency response  $F(i\omega)$  in the complex plane. It is a semicircle at low frequencies (small values of the real part of the pyroelectric current  $I_p$ ) and asymptotic to  $b\pi/2$  at high frequencies (large positive values of the real part of  $I_p$ ). At very high and very low frequencies, Eq. (22) can be expressed by fractional power laws, but with different exponents.

## Thermal Problem

### Radiation Loss, Relaxation Time, and Amplitude of Temperature Oscillation

Radiation losses are determined by the Stefan-Boltzmann law. The power radiated from a surface can be approximated linearly because the heat input is very small:

$$P = \varepsilon\sigma_{SB}(T^4 - T_0^4) \approx 4\varepsilon\sigma_{SB}T_0^3\Theta = G_1\Theta. \quad (23)$$

Here,  $\varepsilon$  is the emissivity of the surface and  $\sigma_{SB}$  the Stefan-Boltzmann constant,  $G$  is the thermal conductance at the interface, and the index 1 indicates the top layer. At room temperature, one gets  $G_1 = 6.12 \text{ W/m}^2 \text{ K}^{-1}$  for  $\varepsilon = 1$ .

In general, heat losses of a slab to the environment are ideally characterized by a single thermal relaxation time

$$\tau_{th} = \frac{cd}{2G}, \quad (24)$$

where  $d$  is the thickness of the slab.

In the case of a semi-infinite solid with a harmonic surface temperature oscillation induced by a heat flux

$$\Phi = \Phi_0 + \Phi_1 \cos(\omega t + \phi), \quad (25)$$

the amplitude of temperature oscillations  $\Theta_{max}$  on the sample surface is given by (Carslaw and Jaeger 1959)

$$\Theta_{\max} \approx \frac{\Phi_1}{\sqrt{\kappa \cdot c \cdot \omega}}. \quad (26)$$

where  $\kappa$  is the thermal conductivity of the irradiated material. Assuming  $\Phi_1 = 1 \text{ W/m}^2$ , the temperature amplitude yields  $\Theta_{\max} = 0.428 \text{ mK}\cdot\text{s}^{-1/2}\cdot(\omega)^{-1/2}$  for low-temperature co-fired ceramics (LTCC) HL2000 (Sea Ceramics Technology 2017a) and  $\Theta_{\max} = 0.571 \text{ m}\cdot\text{K s}^{-1/2}\cdot(\omega)^{-1/2}$  for PZT, respectively.

## Boundary Conditions

Boundary conditions are a set of additional constraints that limit the possible solutions of the differential Eq. (1). Since Eq. (1) is of second order in the spatial coordinates and of first order in time, two boundary conditions for each of the coordinates and one initial condition at  $t = 0$  are necessary. For heat transfer problems, the usually arising surface conditions are (Carslaw and Jaeger 1959):

1. Constant surface temperature at the interface:

$$\Theta(t) = \Theta_s. \quad (27)$$

However, in practice, the surface temperature is often a function of both time and coordinates.

2. No thermal flux across interface:

$$\frac{\partial \Theta}{\partial z} = 0. \quad (28)$$

3. Linear thermal flux at the interface:

$$\kappa \frac{\partial \Theta}{\partial z} + G\Theta = 0. \quad (29)$$

For a thermally isolated surface ( $G = 0$ ), Eq. (29) reduces to Eq. (28).

4. Ideal thermal contact at an interface with different thermal conductivity:

$$\begin{aligned} \kappa_j \frac{\partial \Theta_j}{\partial z} &= \kappa_{j+1} \frac{\partial \Theta_{j+1}}{\partial z}. \\ \Theta_j &= \Theta_{j+1}, \end{aligned} \quad (30)$$

5. Thermal contact with a well-stirred fluid, i.e., with homogeneous fluid temperature:

$$\kappa \frac{\partial \Theta}{\partial z} + \frac{c_f V_f}{A_f} \frac{\partial \Theta}{\partial t} = Q. \quad (31)$$

where  $c_f$  is the volumetric specific heat,  $V_f$  the volume of the fluid, and  $A_f$  the contact area to the fluid. This boundary condition is also used to include the area-specific thermal capacitance (thermal mass)  $c_e d_e$  of metal electrodes (Ploss and Bianzano 1994; Bloß et al. 2000).

## Multilayer Structures

Since the parabolic heat diffusion Eq. (1) has the same structure as the equation of an electrical  $RC$  transmission line, multilayer problems are best treated by the matrix methods commonly used in electrical engineering (Adby 1980). In our case, temperature plays the role of voltage and thermal flux the role of electrical current. Each point is characterized by two quantities – the temperature  $\Theta$  and the heat flux  $\Phi$ . Denoting  $\Theta$  and  $\Phi$  as the temperature and the thermal flux of a slab at  $z = 0$  and  $\Theta'$  and  $\Phi'$  as their values at  $z = d$ , the following relations are obtained (Carslaw and Jaeger 1959):

$$\begin{aligned} \Theta' &= L \cdot \Theta + M \cdot \Phi \\ \Phi' &= N \cdot \Theta + O \cdot \Phi, \end{aligned} \quad (32)$$

with

$$\begin{aligned} L &= \cosh(kd), & M &= -(\kappa k)^{-1} \sinh(kd), & N &= -\kappa k \sinh(kd), \\ O &= \cosh(kd), \end{aligned} \quad (33)$$

and

$$k = \sqrt{\frac{i\omega}{D}}. \quad (34)$$

This may be regarded as a matrix equation

$$\begin{bmatrix} \Theta' \\ \Phi' \end{bmatrix} = \begin{bmatrix} L & M \\ N & O \end{bmatrix} \begin{bmatrix} \Theta \\ \Phi \end{bmatrix}. \quad (35)$$

Here,  $\Theta$  and  $\Phi$  must be multiplied by their time-dependent term  $\exp(i\omega t)$  which is omitted here entirely. It should be again included when real or

imaginary parts have to be taken at the end of the calculation. Assuming perfect thermal contact between the faces of the slabs comprising  $n$  sublayers, the  $r$ th being of thickness  $d_r$ , thermal conductivity  $\kappa_r$ , and thermal diffusivity  $D_r$  with  $\Theta_r$ ,  $\Phi_r$ ,  $\Theta'_r$ , and  $\Phi'_r$  at its left-hand and right-hand faces, respectively, recursive application of Eq. (35) gives for a one-dimensional heat flow across several sublayers,

$$\begin{bmatrix} \Theta'_n \\ \Phi'_n \end{bmatrix} = \begin{bmatrix} L_n & M_n \\ N_n & O_n \end{bmatrix} \begin{bmatrix} L_{n-1} & M_{n-1} \\ N_{n-1} & O_{n-1} \end{bmatrix} \cdots \begin{bmatrix} L_1 & M_1 \\ N_1 & O_1 \end{bmatrix} \begin{bmatrix} \Theta_1 \\ \Phi_1 \end{bmatrix}. \quad (36)$$

Given any two of  $\Theta_r$  and  $\Phi_r$ ,  $\Theta'_r$  and  $\Phi'_r$ , the other two can be found using this method.

The multiplication of matrices in Eq. (36) can be carried out successively. However, deriving explicit formulas for slabs of  $n$  layers is very cumbersome (Blevin and Geist 1974; Zajosz 1979; Samoilov and Yoon 1998). On the other hand, the present method allows numerical values for special cases to be studied very easily by multiplying the corresponding matrices.

In the presence of thermal contact resistances between the slabs or at the surfaces, they may be expressed also in matrix notation and included into Eq. (36) (Carslaw and Jaeger 1959). For example, in the case of a contact resistance  $R_{th,c1}$  between the first and second slab, one finds

$$\Phi'_1 = \Phi_2 = \frac{(\Theta'_1 - \Theta_2)}{R_{th,c1}}, \quad (37)$$

and, correspondingly:

$$\begin{bmatrix} \Theta'_2 \\ \Phi'_2 \end{bmatrix} = \begin{bmatrix} L_2 & M_2 \\ N_2 & O_2 \end{bmatrix} \begin{bmatrix} 1 & -R_{th,c1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} L_1 & M_1 \\ N_1 & O_1 \end{bmatrix} \begin{bmatrix} \Theta_1 \\ \Phi_1 \end{bmatrix}. \quad (38)$$

The final results of this calculation are linear relations between temperatures and fluxes  $\Theta_1$ ,  $\Theta'_n$ ,  $\Phi_1$ ,  $\Phi'_n$  at the two surfaces of the composite slab. The surface conditions provide two more equations, so that all four quantities  $\Theta_1$ ,  $\Theta'_n$ ,  $\Phi_1$ ,  $\Phi'_n$  can be determined.

## Lumped Model

When uniform heating prevails, a simplified lumped model may be applied by solving a bolometer equation (Putley 1970):

$$C_{th} \frac{d\Theta}{dt} + \frac{1}{R_{th}} \Theta = \Phi(t), \quad (39)$$

where  $R_{th}$  is the thermal resistance of the interfaces and  $C_{th}$  the thermal capacitance of the pyroelectric material. For a sinusoidal heat flux, Eq. (25), the steady-state solution of Eq. (39) yields for a thermally coupled slab:

$$\Theta(\omega, t) = \frac{\Phi_1}{cd} \frac{\tau_{th}}{1 + i\omega\tau_{th}} \exp(i\omega t), \quad (40)$$

where  $\tau_{th} = R_{th}C_{th}$  is the thermal relaxation time also given by Eq. (24).

## Thermal Pulse

The thermal problem of an infinite slab in a region  $0 < z < d$  at initial temperature  $\Theta(z,0)$  with linear flux with equal  $G$  at both faces was considered in Carslaw and Jaeger (1959). In the case  $\Theta(z,0) = \Theta_\infty\delta(z=0)$ , the solution is given by:

$$\Theta_\delta(z, t) = \Theta_\infty \cdot \sum_{n=1}^{\infty} C_n \cdot \varphi_n(z) \cdot \exp\left(-\frac{R_n^2 D}{d^2} \cdot t\right), \quad (41)$$

where  $\delta(z)$  is the Dirac delta function and  $\Theta_\infty = \Phi_1 \cdot d / \kappa$  the average temperature of the slab in the short time limit (same as the asymptotic value of  $\Theta(t,z)$  at  $t \rightarrow \infty$  if it is thermally isolated (for a given pulse energy  $Q$  absorbed by the volume of a slab with volumetric specific heat  $c$ , the relation  $Q = c \cdot V \cdot \Delta T$  yields  $\Theta_\infty = Q''/cd$  with  $Q''$  the incident pulse energy per area in  $J/m^2$ ), cf. Eq. (47)),

$$C_n = \frac{2R_n^2 r^2}{1 + 2r + R_n^2 r^2}, \quad (42)$$

$$r = \frac{\kappa}{Gd}, \quad (43)$$

and

$$\varphi_n(z) = \cos\left(\frac{R_n z}{d}\right) + \frac{1}{R_n q} \cdot \sin\left(\frac{R_n z}{d}\right). \quad (44)$$

The roots  $R_n$  have to be determined from the equation:

$$\tan \xi = \frac{2r\xi}{r^2\xi^2 - 1}. \quad (45)$$

For  $r \rightarrow \infty$  ( $G \rightarrow 0$ , thermally isolated slab, i.e., no heat losses) one gets  $R_n \rightarrow n\pi$ . Here, a term

$$\int_0^d \delta(0) dz = 1 \quad (46)$$

has to be added to the sum in Eq. (41) resulting in:

$$\Theta_\delta(z,t) = \Theta_\infty \cdot \left[ 1 + 2 \sum_{n=1}^{\infty} \cos\left(\frac{n\pi z}{d}\right) \cdot \exp\left(-\frac{n^2\pi^2 D}{d^2} \cdot t\right) \right]. \quad (47)$$

This is the solution of Eq. (1) also given in Camia (1967). The value

$$\tau_t = \frac{d^2}{\pi^2 D} \quad (48)$$

denotes the thermal transit time during which the sample reaches internal thermal equilibrium. An approximate solution of Eq. (1),

$$\Theta_\delta(z,t) = \Theta_\infty \cdot \left[ \exp\left(-\frac{t}{\tau_{th}}\right) + 2 \sum_{n=1}^{\infty} \cos\left(\frac{n\pi z}{d}\right) \cdot \exp\left(-\frac{n^2 t}{\tau_t}\right) \right], \quad (49)$$

was proposed in Bauer and Ploss (1988). It is valid for small-enough losses, i.e., large thermal relaxation times  $\tau_{th} \gg \tau_t$ . Here, the thermal relaxation time  $\tau_{th}$  is given by Eq. (24).

For  $r \rightarrow 0$  ( $G \rightarrow \infty$ , thermally coupled to the environment at the back end of the slab), Eq. (45) yields  $R_n \rightarrow (n + 1/2)\pi$ . In this limiting case, Eq. (41) corresponds to the solutions of the thermal pulse problem given in Camia (1967). The solution of the thermal pulse problem with linear flux boundary conditions, Eq. (29), and a homogeneous temperature of the top electrode according to Eq. (31) is given in Bloß et al. (2000). Here, the value  $\langle \Theta \rangle$  relevant for pyroelectric response (cf. Eq. (15)) reaches its maximum with some retardation due to the thermal mass of the electrode. At longer times, the temperature decreases exponentially to room temperature in the same way as predicted by Eq. (41). However, in practice, the pyroelectric response up to the maximum  $\langle \Theta \rangle$  is determined by the thermal pulse shape (Eydam et al. 2016a). In accordance with Duhamel's theorem (Carslaw and Jaeger 1959),  $\Theta(z,t)$  can then be represented by the temperature rise  $\Theta_\delta$  caused by an instantaneous unit pulse input yielding

$$\Theta(z,t) = \int_0^t F(t') \cdot \Theta_\delta(z, t - t') dt', \quad (50)$$

where  $F(t)$  is the normalized pulse intensity.

For engineering applications, approximate solutions of Eq. (1) simplify deconvolution of the measured pyroelectric current or voltage signals. Assuming a cubic depth distribution of temperature, parametric temperature distribution for a heat pulse is given by Leal Ferreira (1989):

$$\Theta = \begin{cases} \frac{2Q''}{cd} \left( 1 - \frac{3z^2}{q^2} + \frac{2z^3}{q^3} \right) & 0 < z < q(t), \\ 0 & z > q(t) \end{cases}, \quad (51)$$

with  $q = 3.51 \cdot (Dt)^{1/2}$  and  $Q''$  the incident pulse energy per area.

## Thermal Steps

A thermal step is given by

$$F(t) = \begin{cases} 0 & t < 0 \\ \Delta\Theta_0 & t \geq 0 \end{cases}. \quad (52)$$

With regard to Eqs. (50) and (52), the solution of Eq. (1) for the case of a thermally coupled back-side, mostly used in practice, has the form:

$$\Theta = \Delta\Theta_0 \left[ \left( 1 - \frac{z}{d} \right) - \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin\left(\frac{n\pi z}{d}\right) \exp\left(-\frac{n^2\pi^2 D}{d^2} t\right) \right]. \quad (53)$$

Here, an approximate temperature distribution is given by

$$\Theta = \begin{cases} \Delta\Theta_0 \left( 1 - \frac{z}{q} \right)^2 & 0 < z < q(t), \\ 0 & z > q(t) \end{cases}, \quad (54)$$

with  $q = 3.36 \cdot (Dt)^{1/2}$ .

## Periodic Surface Temperature

First, a sinusoidally modulated, i.e., in the form of Eq. (25), heat energy  $Q''$  is considered that will be absorbed at the surface. Inserting  $Q''$  into Eq. (50) and taking the approximate Dirac pulse solution  $\Theta_\delta(z,t)$  given by Eq. (49), one arrives at Eq. (55) (Bauer and Ploss 1990): Motivated by its mathematical convenience, a complex exponential is often used for plane wave description as introduced by Landau (Landau and Lifshitz 1969). In this case, only the real part of the complex solution has a physical meaning, provided that all mathematical operations are linear (Carslaw and Jaeger 1959)). In the case of a harmonic temperature excitation, the solution of the thermal problem is often written in complex hyperbolic functions (cf. for instance, (van der Ziel 1973; Lang and Das-Gupta 1986)). Nevertheless, the series solution Eq. (55) may be easily transformed into a hyperbolic function (Prudnikov et al. 1992).

$$\Theta(z,t) = \frac{Q''_0 \tau_{th}}{cd} + \frac{Q''_1}{cd} \Re \left\{ \left[ \frac{\tau_{th}}{1 + i\omega\tau_{th}} + 2 \cdot \sum_{n=1}^{\infty} \cos\left(\frac{n\pi \cdot z}{d_f}\right) \frac{\tau_d/n^2}{1 + i\omega\tau_d/n^2} \right] \cdot \exp(i\omega t) \right\}. \quad (55)$$

A train of rectangular pulses is given by

$$f(t) = \begin{cases} 1 & nT \leq t < nT + \tau_p \\ 0 & nT + \tau_p \leq t < (n+1)T \end{cases}, \quad (56)$$

where  $T = 2\pi/\omega_0$  is the pulse period. Integration of Eq. (25) using a Dirac comb within the pulse length  $[-\tau_p/2, \tau_p/2]$  yields the equation:

$$\Phi(t) = \Phi_0 \frac{\tau_p}{T} \left[ 1 + 2 \sum_{m=1}^{\infty} \frac{\sin(m\omega_0\tau_p/2)}{m\omega_0\tau_p/2} \exp(im\omega_0 t) \right]. \quad (57)$$

For a thermally thick and thermally insulated slab possessing a top electrode with negligible thermal mass, the periodic part of the temperature change yields:

$$\Theta_{\omega_0}(z,t) = \frac{2\Phi_0}{\kappa} \frac{\tau_p}{T} \cdot \Re \left\{ \sum_{n=1}^{\infty} \frac{\sin(m\omega_0\tau_p/2)}{m\omega_0\tau_p/2} \cdot \frac{\cosh \left[ \frac{(1+i)(d-z)}{d_D(n\omega_0)} \right] \cdot \exp(in\omega_0 t)}{\frac{(1+i)}{d_D(n\omega_0)} \cosh \left[ \frac{(1+i)d}{d_D(n\omega_0)} \right]} \right\}, \quad (58)$$

where  $d_D(n\omega_0)$  is the penetration depth of temperature oscillations, Eq. (3), for  $\omega = n\omega_0$ .

Equation (58) illustrates that by traveling through the top layer of the module packaging, higher harmonics of a square wave disappear and the temperature oscillation gradually becomes sinusoidal. Also a phase shift appears as is known for thermal sine waves (Carslaw and Jaeger 1959). Therefore, the following analyzes the thermal problem of a harmonically heated piezoelectric slab exhibiting heat losses to the environment characterized by a single thermal relaxation time or by a distribution of relaxation times.

A periodic triangular-shaped temperature variation may be represented as a sum of sinusoidal changes yielding

$$\Theta(z,t) = \frac{8\Delta\Theta_0}{\pi^2} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2} \exp \left[ -\sqrt{\frac{(2n+1)\omega}{2D}} z \right] \cdot \sin \left[ (2n+1)\omega t - \sqrt{\frac{(2n+1)\omega}{2D}} z \right]. \quad (59)$$

---

## Thermal Methods

The pioneering study of temperature oscillations dates back to 1863 (Angström 1863). In 1956, a technique for pyroelectric coefficient determination in BaTiO<sub>3</sub> single crystals was suggested subjected to periodic steplike heating by flashes of

light (Chynoweth 1956). The application of thermal methods to characterize the polarization homogeneity of pyroelectric films was initiated in 1974 (Peterson et al. 1974). Applying sinusoidally modulated laser radiation, the frequency response of both-side electrode 12- $\mu\text{m}$ -thick PVF and 6- $\mu\text{m}$ -thick PVF<sub>2</sub> films was compared with a model of films with an exponentially decaying polarization which exhibit thermal losses to the environment. Step-function optical or electrical heat inputs were proposed to recast polarization inhomogeneity in the time domain. Since that time, temperature oscillation techniques for nondestructive probing of electric field, space charge, and polarization profiles in dielectric materials are widely used in high-resolution measurements (Collins 1977; von Seggern 1978; Mopsik and DeReggi 1982; Bezdetny et al. 1984; von Seggern et al. 1984; Toureille and Reboul 1988; Lang 1990, 1991, 1998, 2006; Bauer and Ploss 1990; Lang et al. 1992; Ploss et al. 1992; Yilmaz et al. 1993; Ploss and Bianzano 1994; Sessler 1997; Ahmed and Srinivas 1997; Bloß et al. 2000; Reboul et al. 2001; Sakai et al. 2002; Suchaneck et al. 2002; Sandner et al. 2002; Reboul and Mady 2004; Notingher et al. 2004; Mellinger 2004; Mellinger et al. 2005a, b, 2006; Dagher et al. 2006; Petre et al. 2006; Malyshkina et al. 2007; Gerlach et al. 2007; Holé 2008; Movchikova et al. 2008a, b, 2009, 2010; Stewart and Cain 2008; Lang and Fleming 2009; Qiu et al. 2010; Malyshkina et al. 2010; Reboul 2011; Suchaneck et al. 2011, 2012, 2013a, b; Zheng et al. 2013; Eydam et al. 2014, 2015, 2016a, b, c; Imburgia et al. 2016).

## Thermal Methods in the Frequency Domain

### Pyroelectric Response of a Harmonically Heated Slab Exhibiting Thermal Losses to the Environment

For the lumped model, Eq. (39), the pyroelectric current yields

$$I_p(\omega) = \frac{\Phi_1 A \cdot p_0}{cd} \frac{i\omega\tau_{th}}{1 + i\omega\tau_{th}} \exp(i\omega t). \quad (60)$$

Following Eq. (55),  $I_p(\omega)$  of a harmonically heated piezoelectric slab exhibiting heat losses to the environment, characterized by a thermal relaxation time  $\tau_{th}$ , takes the form (Bauer and Ploss 1990):

$$I(\omega) = \frac{\Phi_1 A}{cd} \left( p_0 \frac{i\omega\tau_{th}}{1 + i\omega\tau_{th}} + \sum_{n=1}^{\infty} p_n \frac{i\omega\tau_t/n^2}{1 + i\omega\tau_t/n^2} \right) \exp(i\omega t), \quad (61)$$

where  $p_0$  is the average pyroelectric coefficient and  $p_n$  are the spatially dependent parts of the pyroelectric coefficient,

$$p_0 = \frac{1}{d} \int_0^d p(z) dz, \quad p_n = \frac{2}{d} \int_0^d p(z) \cos\left(\frac{n\pi z}{d}\right) dz. \quad (62)$$

For homogeneous polarization of the slab, one arrives at Eq. (60). For a continuous distribution of relaxation times, Eq. (60) can be generalized by introducing the Havriliak-Negami function (Suchaneck et al. 2013b):

$$I_p(\omega) = \frac{\Phi_1 A}{cd} p_0 \left[ 1 - \frac{1}{[1 + (i\omega\tau_{th})^a]^b} \right]. \quad (63)$$

For  $a = 1$  and  $b = 1$ , Eq. (63) reduces to Eq. (60). Two special cases are the Cole-Cole relaxation for  $a < 1$  and  $b = 1$  as well as the Cole-Davidson relaxation for  $a = 1$  and  $b < 1$ . The average charge induced due to the pyroelectric effect depends on the mean temperature and is given by

$$\langle Q \rangle = \langle p\Theta \rangle \approx \frac{p_0}{d} \Re \int_0^d \Theta(z) \exp(i\omega t) dz. \quad (64)$$

Considering Eqs. (55 and 61), the spatial polarization distribution provides a small contribution to the imaginary part of the pyroelectric current in the frequency range of  $1/\tau_{th} < \omega < 1/\tau_d$  given by (Suchaneck et al. 2012):

$$I_p(\omega)_2 = \frac{\Phi_0 A}{c \cdot d} \sum_{n=1}^{\infty} p_n \frac{i\omega\tau_t/n^2}{1 + i\omega\tau_t/n^2} \approx i\omega \frac{\Phi_0 A}{c \cdot d} \sum_{n=1}^{\infty} \frac{p_n \tau_t}{n^2}. \quad (65)$$

Since  $\cos(n\pi z/d) < 1$ , one finds

$$\Im I_p(\omega)_2 < 2 \frac{\Phi_0 A}{c \cdot d} \omega \tau \cdot p_0 \cdot \sum_{n=1}^{\infty} \frac{1}{n^2} = 2 \frac{\Phi_0 A}{c \cdot d} \omega \tau \cdot p_0 \cdot \zeta(2) = \frac{\pi^2}{3} \frac{\Phi_0 A}{c \cdot d} \omega \tau \cdot p_0, \quad (66)$$

where  $\zeta$  is the Riemann Zeta function. No information about the polarization profile can be extracted within a frequency region determined by

$$\frac{\pi^2}{3} \omega \tau \ll \frac{\omega \tau_{th}}{1 + (\omega \tau_{th})^2} \approx \frac{1}{\omega \tau_{th}}. \quad (67)$$

The frequency response of Eq. (60)

$$\Re F = \frac{\Re I_p(\omega)}{(\Phi_1 A p_0 / cd)} = \frac{(\omega \tau_{th})^2}{1 + (\omega \tau_{th})^2}, \quad \Im F = \frac{\Im I_p(\omega)}{(\Phi_1 A p_0 / cd)} = \frac{\omega \tau_{th}}{1 + (\omega \tau_{th})^2}, \quad (68)$$

represents a semicircle in Nyquist complex plane mapping. The maximum of the imaginary part of the pyroelectric current allows an estimation of the thermal relaxation time, Eq. (24), of the slab.

### The Laser Intensity Modulation Method (LIMM)

In 1986, a technique for the determination of the spatial distribution of polarization in polymers in the frequency domain was introduced (Lang and Das-Gupta 1986) which is an extension of the thermal pulse method (Collins 1977) and a technique for dynamic pyroelectric measurements (Chynoweth 1956). In the so-called laser intensity modulation method (LIMM), a modulated laser beam is absorbed by the top electrode and generates a temperature oscillation into the film which propagates in the sample and causes local changes in temperature, thermoelastic strains, pyroelectric response, etc. LIMM determines the depth profile of the pyroelectric coefficient from the pyroelectric current spectrum caused by the interaction of the temperature oscillations and the unknown polarization distribution since the penetration depth of the temperature oscillations, Eq. (3), depends on the modulation frequency. Here, the temperature distribution produced along the thickness of the sample will be non-uniform, i.e., the laser modulation frequency is  $f > D/\pi d^2$ . A similar technique using non-coherent light was developed and applied to ferroelectrics about 2 years earlier in Russia (Bezdetyn et al. 1984). The authors considered a lossy ferroelectric, derived a Fredholm integral equation similar to that of the LIMM problem, and used sinusoidally modulated non-coherent light at frequencies up to 1 kHz for measurement.

In Devonshire's phenomenological theory of ferroelectricity, a polynomial thermodynamic potential expressed in terms of deviation from the non-polar state is considered and a linear temperature dependence in the first coefficient is introduced to account for the temperature dependence of dielectric susceptibility. Higher-order coefficients are assumed to be independent of temperature (Devonshire 1954). In this case, the pyroelectric voltage response  $p/e$  is simply related to the spontaneous polarization  $P_s$  (Liu and Zook 1974):

$$\frac{p}{e} = \frac{P_s}{C}, \quad (69)$$

where  $C$  is the Curie constant. Therefore, LIMM is often denoted as a technique for polarization profiling.

In the frequency domain, the pyroelectric current, Eq. (12), possesses a complex amplitude  $I_\omega$ :

$$I_p(\omega) = I_\omega \exp(i\omega t) = \frac{A}{d} \int_0^d r(z) \frac{\partial}{\partial t} \Theta_\omega(z,t) dz, \quad (70)$$

where  $\Theta_\omega(z,t)$  is the periodic part of the solution of the thermal problem. According to Eqs. (55 and 70), the sample depth region that is thermally probed most efficiently becomes small at high frequencies and is located in the vicinity of the penetration depth  $d_D$  of the thermal wave, Eq. (3) (Bauer and Bauer-Gogonea 2003).

## Spatial Resolution

Initially, the spatial resolution of LIMM was estimated using simulations for Dirac delta function polarization distributions (Lang 1991). The pyroelectric current corresponding to this polarization distribution at a given depth within the sample was calculated. The simulated current spectrum superimposed with white noise was transformed back to a polarization distribution. The back-transformed polarization distributions were not very sensitive to the amount of noise being in the 0.5–10% range. Their full width at half-maximum (FWHM) amplitude was then considered as a measure of minimum resolvable distance yielding

$$FWHM \propto z^{0.4}. \quad (71)$$

Numerical simulations based on this approach have shown that the FWHM increases almost linearly up to a depth  $z/d \approx 0.7$  from the harmonically heated surface (Sandner 2003). It increases with increase of the square root of noise. Above  $z/d \approx 0.7$ , the FWHM deteriorates, going through a minimum at  $z/d \approx 0.83$  and an overall maximum at  $z/d \approx 0.9$ . Near the heated surface, the resolution limit was  $\delta z = 0.25 d_D$ . For PZT thin films, characterized with a system bandwidth of  $f_{\max} = 2$  MHz, the near-surface limit of the minimum resolvable distance amounted to  $z_{\min} = 65$  nm.

Spatial resolution refers to the wave vector of spatial frequencies corresponding to the minimum separation between adjacent points that can be both detected and interpreted. For the Fourier transform of an image, the highest spatial frequency present in the resulting Fourier-domain image represents the information limit. For an analog signal subjected to an analog-to-digital conversion with a bandwidth  $f_{\max}$ , the sampling rate according to Nyquist's sampling theorem should be at least  $2f_{\max}$ . Thus, the minimum resolvable distance is given by

$$\delta z = \frac{v}{2f_{\max}} = 2.506 \sqrt{\frac{D}{2f_{\max}}} = 2.506\sqrt{Dt}, \quad (72)$$

with  $v$  the diffusion velocity, Eq. (5).

Considering the position at 50% of the amplitude, one finds:

$$\delta z = 2\ln(2)d_D = 1.106\sqrt{Dt}. \quad (73)$$

The minimum resolvable distance was also defined as the distance needed for the thermal oscillation to go from 90% to 10% of its amplitude (Dagher et al. 2006) where

$$\delta z = 1.75\sqrt{Dt}. \quad (74)$$

Thus, the spatial resolution depends essentially on the heat diffusivity of the studied material and on the bandwidth square root. For materials with a larger thermal diffusivity, the spatial resolution is lower in comparison to a

polymeric material. The diffusive nature of heat transport results in a strongly nonuniform spatial resolution across the sample. A high resolution (small resolvable distance) is obtained in the vicinity of the thermally excited electrodes. It rapidly decreases toward the middle of the sample owing to the physics of thermal diffusion. LIMM data lose any information on the spatial resolution far away from the thermally excited electrode (Lang 1998, 2004; Bauer and Bauer-Gogonea 2003).

Considering a planar uniform sample perfectly matched at its back side, a minimum resolvable distance was estimated for different measuring techniques (thermal pulse method, thermal step method, LIMM) in Holé (2008) who took two kinds of spatial resolution into account. The first one described the position accuracy of a detected charge; the second one was related to the discernment of two distinct identical charges. The latter defines the resolving power which is not considered in this chapter since polarization, except in domain walls and very near electrical contacts, is a slowly changing property. Exactly for that reason, the polarization depth profile is subjected to the principle of parsimony (choosing the smoothest profile based on physical intuition) (Lang 2006). LIMM usually collects data from equal measurement repetitions at each frequency. In this case, the signal-to-noise ratio (SNR) defined as signal amplitude to random means square (RMS) noise is constant throughout the sample and the minimum resolvable distance is given by

$$\delta z = \frac{2.72z}{\sqrt{SNR}}. \quad (75)$$

The Rose criterion (Rose 1973) states that an SNR of at least 5 is needed to be able to distinguish image features at 100% certainty. Thus, one arrives to  $\delta z < 1.22z$ . In Sandner (2003), LIMM was applied to PZT thin films by collecting 40 measurement repetitions at each frequency. This led to a relative noise amplitude of 1.3 ... 1.9% and, correspondingly,  $\delta z = 0.310z \dots 0.375z$ . The slope of this linear dependence was in good agreement with numerical simulations yielding about 0.25 for a relative noise amplitude in the range of 1–5%.

Conventional LIMM measurements are carried out over a range of frequencies, typically from a few Hz to about 500 kHz which brings the minimum resolvable distance to about 2.5 μm for a film thickness of 25 μm (Ahmed and Srinivas 1997). In order to obtain a higher spatial resolution, the measuring system bandwidth must be as high as possible. Conventional equipment is limited to a few GHz providing a minimum resolvable distance as high as a few 10 nm. The sub-micron resolution of LIMM near the thermally-excited electrode was first demonstrated in (Ploss and Bianzano 1994). However, in high-frequency LIMM measurements, the linear flux boundary conditions at the top electrode, Eq. (29), must be complemented by a condition of homogeneous top electrode temperature, Eq. (31), to account for the impact of the electrode thermal mass on the pyroelectric current amplitude and phase at frequencies exceeding 1 MHz.

Since LIMM is better suited for polarization determination near the surface, a modification of LIMM was proposed distinguished by much higher range of modulation frequencies and analysis of data from only one side of the sample called surface LIMM (Lang et al. 1992).

### Analysis of LIMM Data

A mathematical deconvolution technique is required to compute the polarization distribution from the pyroelectric current spectrum of Eq. (70). However, Eqs. (12 and 70) are Fredholm integral equations of the first kind which are known to represent a mathematically ill-posed problem (Groetsch 2007), i.e., any arbitrary small noise contribution in the pyroelectric current or inaccuracies due to discretization during digital signal processing makes it possible to find different pyroelectric distributions with random deviations. The physical reason for the ill-posed nature of Eq. (70) is the nonuniform spatial resolution.

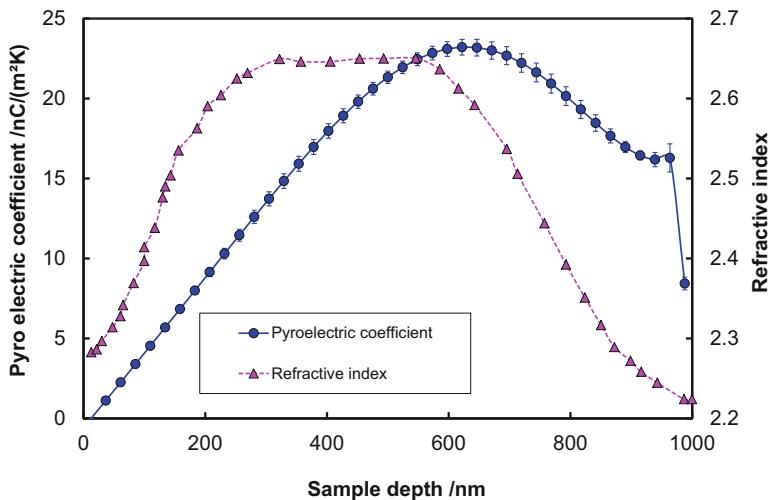
There are various approaches to solve such ill-posed problems (Lang 2004; Mellinger 2004; Lang and Fleming 2009). A more sophisticated approach to filter out high-frequency oscillating solutions is given by the Tikhonov regularization method (Groetsch 2007) where additional a priori information about the desired solution is incorporated. The regularization method utilizes the physical requirement that the polarization distribution must be smooth. In detail, the application of Tikhonov regularization to LIMM is described in Sandner et al. (2002).

Considering either a thermally isolated or thermally coupled one-layer system, Eq. (70) may be solved analytically by means of the Mellin transform (Sandner et al. 2002). The result is an integral in the complex plane which may be evaluated numerically using standard mathematics software packages.

Figure 3 compares the polarization profile of a PZT thin film obtained by LIMM and applying Tikhonov regularization (Sandner et al. 2002) with the refractive index profile of the same film determined by multi-angle spectroscopic ellipsometry (Suchaneck et al. 2002). The refractive index and pyroelectric coefficient profiles differ due to different boundary conditions. The refractive index profile is symmetric due to open-circuit mode during measurement, while the pyroelectric response is determined in a short-circuit mode. It becomes asymmetric due to current flow in the external circuit.

### Two-Dimensional and Three-Dimensional LIMM

For more than half a century, the distribution of electric polarization at crystal surfaces has been displayed by means of pyroelectric effects when a heat spot produced by a sharply focused electron beam was scanned across it (Burfoot and Latham 1963; Hadni et al. 1965). Later, the electron beam was replaced by a 50- $\mu\text{m}$ -diameter optical source for studying antiparallel polarization distributions in opaquely electroded single crystals of barium titanate (Clay et al. 1974). The smallest region of homogeneous polarization detectable by the probe was approximately equal to the effective source size. A helium-neon laser with a spot diameter  $R = 60 \mu\text{m}$  square-wave modulated at 16 and 800 Hz was used for TGS pyroelectric mapping to evaluate spontaneous polarization and domain nucleation sites in Hadni



**Fig. 3** Comparison of the pyroelectric coefficient depth profile of a PZT thin film with a thickness of 1  $\mu\text{m}$  (dots) and its refractive index profile (triangles)

and Thomas (1972). A scanning pyroelectric microscope comprising a diode laser at 780 nm, focused through a glass substrate, was reported in Yilmaz et al. (1993). Scans of pyroelectric response of layers of a nonlinear optical side-chain polymer, periodically poled with a period of 110  $\mu\text{m}$ , were performed.

The two-dimensional (2D) scanning laser intensity modulation method was developed to detect various types of defects in PZT ceramic materials including localized sample depoling induced by thermal and mechanical treatments and by environmental degradation under high DC bias (Stewart and Cain 2008). The laser beam diameter was about 50  $\mu\text{m}$ . Lateral thermal diffusion in the top Ni electrode increases the effective beam diameter at lower frequencies. Being 50  $\mu\text{m}$  at a frequency of 10 kHz, it becomes about 180  $\mu\text{m}$  at 100 Hz and approximately 500  $\mu\text{m}$  at 1 Hz. In order to give a good signal-to-noise ratio, a sample scan was performed at a single fixed frequency in the range of 1 Hz to 100 kHz. Below 1 Hz, the control of the thermal environment is complicated and data acquisition times are extended. Higher frequencies produced noisier data as the current-to-voltage amplifiers bandwidth cutoff frequency was approached. For the FEMTO DLPCA-200 variable gain current amplifier with a capacitive load of 30 nF used in own experiments, the signal power at a gain of  $10^{10}$  V/A is attenuated by 50% (or 3.01 dB) at about 500 Hz. Without a capacitive load, the cut-off frequency is 1130 Hz. Images for PZT were recorded for  $f = 97$  Hz corresponding to a penetration depth  $d_D = 38 \mu\text{m}$  (cf. Eq. (3)). The time at each point was 470 ms, leading to image acquisition times of between 9 and 32 h. With a step size of 2  $\mu\text{m}$ , the smallest distinguishable feature was of the order of 30  $\mu\text{m}$ . When the step size of the scan is larger than the range of thermal diffusion zone under the beam, a sharp drop in temperature occurs at the old position and a sharp rise at the new one leading to an

additional pyroelectric current which cancels out the wanted signal. Since the traveling  $x$ - $y$  translation stage can introduce acoustic noise into the measurements, particularly when examining piezoelectric ceramic samples with a relatively high mechanical quality factor, galvanometer mirror scanners were used in Stewart and Cain (2008). This leads to an improvement in the speed of the system by a factor of 5. An image consisting of  $80 \times 80$  points required approximately 30 min for acquisition at a step size of 50  $\mu\text{m}$ . Piezoelectric fibers with a diameter of 250  $\mu\text{m}$  were imaged. They were embedded in an epoxy matrix as artificial defects. Also, soldered regions exhibiting a good thermal and mechanical coupling to the surroundings were displayed.

### Alternative Thermal Wave Method (ATWM)

In 2001, the ATWM was proposed for measurements on thin dielectrics located in power capacitors (Reboul et al. 2001) which is based on Eq. (59). The main difference to the thermal step method (cf. section “Thermal Step Method (TSM)”) is lower noise and a better resolution by using a continuous thermal excitation rather than a single stimulus (Reboul and Mady 2004). Similar to the thermal square wave method (cf. section “Thermal Square Wave Method (TSWM)”), the periodic pyroelectric signal consists of higher harmonics enabling depth profiling. Note that the triangular function is the convolution of two identical unit rectangular functions. Since the Fourier series of a triangular function contains only odd harmonics, the phase shift of the pyroelectric current should be measured more easily. However, in practice, it is quite cumbersome to realize a perfect triangular temperature excitation (Reboul et al. 2001). In 2011, the ATWM was improved in terms of accuracy by applying two simultaneous thermal waves at both surfaces of the specimen (Reboul 2011). The tuning of the phase lag of the two thermal excitations allows easy generation of different temperature distributions in the sample. Hot or cold zones in the sample can be matched depending only on the phase shift. A kind of stationary thermal wave is created. Therefore, each slice in the sample is thermally probed in a different manner using different phases, especially in the middle of the thickness where thermal interferences phenomena are the most effective. However, the sample thickness must be smaller than half of the thermal wavelength to ensure thermal interferences inside, e.g., a quarter of wavelength.

## Thermal Methods in the Time Domain

### Thermal Pulse Method (TP)

Thermal pulse methods date back to the description of the laser flash method for the characterization of thermo-physical properties of thin solid materials in 1961 (Parker et al. 1961).

In the presence of an internal electric field or a polarization, the thermal pulse gives rise to a time-dependent pyroelectric current in accordance with Eq. (12). The thermal pulse technique is advantageous in terms of acquisition speed of data. It provides a larger pyroelectric signal within a shorter measuring time (factor of 50) if

thermal pulse trains are applied which are generated by a pulsed laser (Mellinger et al. 2005a). Another benefit is the minimization of temporal changes of the polarization or space charge profile. Both may change over the course of the measurement. There are two limiting times within which the desired information can be extracted: the thermal transit time during which the sample reaches internal thermal equilibrium, Eq. (48), and the thermal time constant of the sample for reaching thermal equilibrium with the environment, Eq. (24).

### Spatial Resolution

With pulsed excitation, time resolution is of primary interest. This requires a short pulse length  $\tau_p$ . Additionally, the thermal transit time  $\tau_t$ , Eq. (48), across the sample should be as short as feasible, i.e., the material under test should be as thin as possible. To ensure accuracy of the measurement setup electronics, the time constant  $\tau_e$  should be larger than any other time constant in the system, i.e.,  $\tau_e \gg \tau_{th} \gg \tau_t$ .

With a pulse length of 60  $\mu\text{s}$ , it is possible to obtain routinely the first 7 to 10 Fourier coefficients of the polarization distribution, enabling a resolution of the order of 2.5  $\mu\text{m}$  in a 25  $\mu\text{m}$  thick film (Mopsik and DeReggi 1982). In Amjadi (1996), the space charge distribution was measured in silicon dioxide of 1  $\mu\text{m}$  thickness by using a short laser pulse (70 ps). Here, the resolution limit estimated as  $\delta z = (2D\tau_p)^{1/2}$  was 11 nm. Very fast temperature variations can be generated by using femtosecond laser pulses (Wu and Zhang 1995; Perrin et al. 1996). The use of an optical instrumentation with laser pulse lengths of the order of femtoseconds (Dagher et al. 2006) brought the resolution to values between 22 nm (in polymeric materials) and 50 nm (in silicon dioxide). However, the signal-to-noise ratio was low, so the method needs further investigations.

Another restriction of the thermal pulse technique is the rapid decay of the spatial resolution with sample thickness (von Seggern 1978). The number of realistic Fourier coefficients and, thus, the spatial resolution decreases linearly with sample depth. Taking the signal-to-noise ratio  $\text{SNR}_0$  at position  $z_0$  for reference, the simplified model of a planar uniform sample thermally coupled at its backside in Holé (2008) predicts an increase of the minimum resolvable distance as the square of the sample depth. On the other hand, the amplitude of  $\Theta$  decreases exponentially (cf. Eq. (47)) leading to a corresponding decrease of the SNR. This causes a further deterioration of spatial resolution at larger sample depths limiting – in combination with the difficult signal deconvolution – the use of the thermal pulse method.

### Analysis of Pulse Data

In the thermal pulse technique, the signal is recorded in the time domain. A problem of the thermal pulse method, which also caused its limited use, is that similar to LIMM, the output signal is subjected to mathematical deconvolution in order to trace back the real electric field or polarization distribution.

Applying a thermal pulse to the surface of a thin dielectric and measuring the voltage response, one could derive the charge distribution of electrets by choosing an assumed discrete charge distribution, calculating the shape of the  $\Delta V(t)$  function that

this distribution generates and iterating it manually until the experimental data agrees with the assumption (Collins 1977). Later, the polarization distribution in poled polymer films was evaluated by means of Fourier analysis of the current response of the thermal pulse experiment (Mopsik and DeReggi 1982). However, only a limited number of Fourier coefficients can be determined up to the terms with  $n$  of order  $N = (\tau_t/\tau_p)^{1/3}$ . Here,  $\tau_t$  is the thermal transit time, Eq. (47), and  $\tau_p$  the pulse length. The minimum resolvable distance of the thermal pulse experiment is then roughly  $d/N$ . It can be increased by collecting data from both sides of the sample, taking advantage of the high near-surface depth resolution. Consequently, Fourier analysis should be useful only for the determination of charge centroids (von Seggern et al. 1984). An approximate method for the analysis of thermal pulse experiments was developed in Bauer (1993). The method is based upon the short-term solution of the thermal problem and upon appropriately constructed differential operators in terms of  $s = (Dt)^{1/2}$ . Similar to Eq. (17), it is well-suited to the study of the surface region of the investigated film.

To take advantage of the high near-surface depth resolution of the thermal pulse technique, the measurement needs to be performed on both sides of the sample (Ploss et al. 1992). Recently, the space charge depth profile in various electret polymers has been measured in both the time and the frequency domain using thermal pulses and waves, respectively (Mellinger et al. 2005a). Here, the time domain signal  $I(t_i)$  was converted to the frequency domain via a discrete fast Fourier transform and divided by the measured transfer function of the measuring setup. Besides easier analysis of the data in the frequency domain, the correction of the frequency-dependent gain and phase shift of the preamplifier is an additional benefit. A way to make the deconvolution procedure easier is the Fourier transform of the transient current and calculation of an approximate profile by means of Eq. (17) (Zheng et al. 2013). For focused laser beam scanning, a two-dimensional “coupled neighbors” model was developed where the current signals from several adjacent beam pointings along a scan line were coupled together (Aryal and Mellinger 2013). A Monte Carlo method was then applied for the deconvolution. This technique was proposed to enhance the lateral resolution by a factor of 2–3.

The approximate temperature distributions given by Eqs. (51) and (54) (Biot 1970) were applied to deconvolute a pyroelectric signal by repeated differentiations of products of signal and position (Leal Ferreira 1989).

### Three-Dimensional Thermal Pulse Method

Recently, the thermal pulse method described in Mellinger et al. (2005a) was extended into three dimensions (Mellinger et al. 2005b). Polarization maps of 11- $\mu\text{m}$ -thick poled PVDF films were recorded at a laser wavelength of 532 nm, a pulse length of about 5 ns, and a pulse energy density of less than 0.1 J/cm<sup>2</sup>. The latter was limited in order to avoid ablation damage of the irradiated electrode. The scan covers an area of  $7 \times 7 \text{ mm}^2$  with a lateral resolution of 200  $\mu\text{m}$ . Acquisition of the 1296 thermal pulse measurements took about 3.5 h and produced 2 GB of data. At a depth of 1  $\mu\text{m}$  below the PVDF surface, the lateral resolution – estimated as the

imaged width of the transition from the poled to the unpoled area – was 38 µm and, hence, slightly larger than the spot size. At a depth of 4 µm, the corresponding value amounted to 105 µm. This reduction in lateral resolution was attributed to the thermal diffusivity of the metal electrode being about 1000 times greater than that of the polymer, so that the heated zone in the metal spreads out about 30 times faster than in the polymer (cf. Eq. (3)). Thus, the lateral resolution at greater depths is limited by thermal diffusion in the electrode. The spatial resolution in the  $z$  direction was less than 0.5 µm.

Later, this technique was named thermal pulse tomography (TPT) and applied to polarization mapping of a piezoelectric P(VDF-TrFE) sensor cable for evaluating the poling process (Mellinger et al. 2006). A lateral resolution of 200 µm was claimed. The required time was about 10 s per beam point. In (Qiu et al. 2010), a honeycomb structure was fabricated by attaching two polycarbonate films to a double-sided adhesive-tape grid cut by a computer-controlled laser. The structure was charged by means of internal dielectric barrier discharges and the negatively charged internal void surfaces were revealed through exposure to xerographic toner and via 3D mapping of the electric field by means of TPT. Both polarization patterns were consistent with the grid structure. This qualifies TPT as a very useful technique for investigating electric field profiles in polymer ferroelectrets.

### Thermal Step Method (TSM)

The thermal step method was proposed for the determination of the space charge density in polyethylene (Tourelle and Reboul 1988). Since the sample was coupled to a heat exchanger, the minimum resolvable distance was low. Nevertheless, features in the order of 1 µm of the polarization distribution in 20-µm-thick fatigued films of ferroelectric vinylidene fluoride/trifluoroethylene copolymer were resolved by this step excitation technique (Sakai et al. 2002). On the other hand, for thin enough samples, the spatial resolution will be limited by the thermal transit time of the sample, Eq. (48), and the rise time of laser intensity. By means of a faster heating of the thermal diffuser using a discharge of several capacitors through a resistor, charge distributions in impregnated paper sheets of 80 µm thickness were investigated with a resolution in the order of a few 10 µm (Agnel et al. 1996). TSM was demonstrated to also provide an estimate of the space charge distribution in the semiconductor, in the oxide and at the oxide/semiconductor interface (Notingher et al. 2004). Here, the sensitivity of the TPM was supposed to be at least one order of magnitude above that of the capacitance-voltage method allowing charge detection in dielectrics that cannot be observed otherwise (Notingher et al. 2012).

The main benefit of the TSM is that it provides better spatial resolution than the thermal pulse method far enough from the surface (Holé 2008).

### Thermal Square Wave Method (TSWM)

TSWM at a single frequency determines the pyroelectric coefficient depth profile and, thus, the polarization distribution of ferroelectrics from the time dependence of

the pyroelectric current at a certain heat pulse modulation frequency. It assumes a constant heating rate during pulse time and a constant wave velocity and consists of constraints related to the penetration depth of the temperature oscillation. At constant heating, changes in the pyroelectric current reflect changes of  $dP_s/dT$  in the primary pyroelectric coefficient (cf. Eq. (7)) (Glass 1968). Since the rectangular function may be represented as a sum of cardinal sine functions, the pyroelectric signal contains information of a broad frequency region limited by the bandwidth of the measurement setup.

Following Eqs. (15 and 58) and taking into account that adiabatic conditions are satisfied by  $T = 2\pi/\omega_0 < d^2/\pi D$  (cf. Eq. (3)), an approximate depth profile of the pyroelectric coefficient yields (Malyshkina et al. 2007):

$$p(z_D) \approx \frac{I_p(t)\omega_0\kappa}{8\pi A\Phi_0} \Re \left\{ \left( \sum_{n=1}^{\infty} \frac{\sin^2(n\omega_0\tau_p/2)}{n\omega_0\tau_p/2} \cdot \frac{i \cdot d_D(n\omega_0)^2}{(1+i)^2 \cdot z_D} \cdot \left[ 1 - \exp\left(-\frac{(1+i)z_D}{d_D(n\omega_0)}\right) \right] \right)^{-1} \right\}, \quad (76)$$

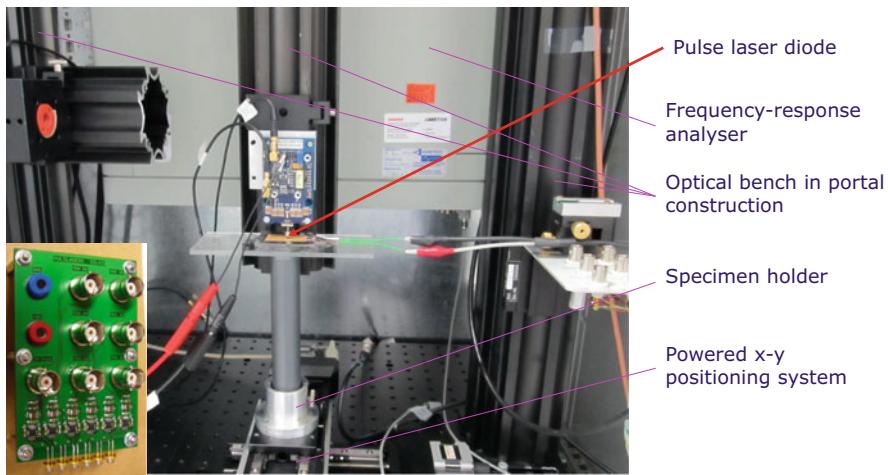
with  $z_D = vt = 2(\pi D/T)^{1/2} \cdot t$  (cf. Eq. (5)). TSWM is then the record of the pyroelectric coefficient depth profile by means of Eq. (76). The time dependence of the pyroelectric current was obtained on both sides of the sample using an IR LED (wavelength  $\lambda = 930\text{--}960$  nm) controlled by a wave function generator and a power amplifier for thermal excitation. The pyroelectric current was converted to voltage by an operational amplifier OP297 with a gain of  $2.5 \cdot 10^8 \text{ V}/\mu\text{A}$  and a bandwidth of 1000 Hz. Signal sampling was performed by a 12-bit analog-digital converter.

TSWM was initially proposed for the characterization of ferroelectric ceramics, e.g., PZT (Malyshkina et al. 2007). It was also demonstrated to record the pyroelectric coefficient profiles of  $\text{BaTi}_{1-x}\text{Sn}_x\text{O}_3$  functional ceramics with a tin gradient of  $0.075 \leq x \leq 0.15$  (Movchikova et al. 2008b), of  $(\text{Sr},\text{Ba})\text{Nb}_2\text{O}_6$  single crystals doped with Rh, Eu, Cr (Movchikova et al. 2008a, 2009), of  $(1-x)\text{Pb}(\text{Mg}_{1/3}\text{Nb}_{2/3})\text{O}_3\text{-xPbTiO}_3$  single crystals (Movchikova et al. 2010), and of  $\text{Sn}_2\text{P}_2\text{S}_6$  films on aluminum substrates (Malyshkina et al. 2010). Similar to other thermal techniques in the time domain, the SNR of the TSWM signal decreases with increasing sample depth yielding a higher resolution near the surface.

## Instrumentation

### Experimental Setup

For LIMM measurements, the samples were periodically heated by an array of six laser diodes or a single laser diode (LCU985041A, Laser Components GmbH, Olching, Germany) both square-wave-modulated with frequencies of up to 1 kHz. Each laser diode had a power of 14 mW at a wavelength of 980 nm. Amplitude and phase of the pyroelectric current were determined by an impedance/gain-phase analyzer (Solartron 1260, Solartron Analytical, Farnborough, UK) with DC



**Fig. 4** Experimental setup for LIMM and thermal pulse measurements. Separately, a laser array is shown which is used for LIMM evaluation of laminar samples

coupling. In order to reduce noise, up to 50 measurement repetitions were used for averaging. The experimental setup is shown in Fig. 4.

Thermal pulse measurements were carried out by heating the samples with a pulsed laser diode (LC905D1S3J09UA, 905D3S3J09S and 905D5S2L3J08, Laser Components GmbH, Olching, Germany) at a wavelength of 905 nm with a peak power of up to 230 W, pulse widths of 0.1 to 0.5  $\mu$ s, and a repetition frequency of 1 Hz. Laser pulse power was measured by means of an 843-R-USB power meter comprising a 919P-003-10 thermopile sensor (Newport Corp., Irvine, CA). The obtained single pulse energy amounted up to 10  $\mu$ J. The pulse shape was evaluated using a HCA-S high-speed current amplifier with integrated silicon avalanche photodiode (Femto Messtechnik, Berlin, Germany). Ten measurement repetitions were averaged. The pyroelectric current was transformed to a voltage by a low-noise current amplifier (SR570, Stanford Research Systems, Sunnyvale, CA) and additionally filtered by a home-built 50 Hz notch filter before recording by a Waverunner<sup>®</sup> Xi-A oscilloscope (LeCroy, Chestnut Ridge, USA) with DC coupling. A total of  $10^6$  data points were stored for data processing.

To compensate for drift in laser intensity, the LIMM spectra may be corrected by performing a calibration with fast photodiodes usually being an internal part of laser diodes. On the other hand, the heat flux which is really absorbed by the sample generating temperature oscillations can only be estimated.

A mean to compensate temporal effects in the measuring setup is working with a random frequency scan rather than a sequential one. To account for the higher resolution of LIMM near the surface, a logarithmic spacing might be adopted for the  $z$  coordinate (Lang 2004; Mellinger 2004; Petre et al. 2006). However, data analysis by means of a discrete Fourier transform requires a fixed coordinate spacing as discretization parameter.

A crucial factor of the measurement setup is the gain/phase characteristics of the preamplifier. They depend on gain, filter time constants, capacitive load, and the modulation depth. Commercial preamplifiers are designed for mainly resistive loads. On the other hand, the capacitances of the piezoelectric samples are in the range of 5 to 170 nF. This reduces significantly the cut-off frequency of the preamplifier given in its specification.

Pulse measurements require an appropriate tradeoff between gain and bandwidth. Setups with large bandwidths (above 20 kHz) and lower gain lead to high-frequency noise while large gains, e.g.  $10^{10}$  V/A and small bandwidths (below 100 Hz), produce significant, unstable DC-offsets which might be avoided by adding a high-pass filter with a cutoff frequency of 1 Hz. Long signal recording times, e.g., 1 s at a bandwidth of 200 Hz, increase significantly the signal-to-noise ratio. By reducing the signal recording time, the information at low frequencies is lost.

The gain-phase bandwidth problem can be solved by combining two data sets, a low-noise mode with substantial signal distortion at high frequencies and a high-bandwidth mode where noise becomes more and more dominant at low frequencies with a smooth transition function (Mellinger et al. 2005a):

$$\tilde{J}_{\text{comb}}(f_n) = \tilde{J}_{\text{hbw}}(f_n)\varphi(f_n) + \tilde{J}_{\text{ln}}(f_n)(1 - \varphi(f_n)), \quad (77)$$

where  $J_{\text{hbw}}(f_n)$  and  $J_{\text{ln}}(f_n)$  are the frequency spectra of the preamplifier signals Fourier transformed and divided by the gain. The transition function was chosen as

$$\varphi(f_n) = \frac{1}{2} \left\{ \tanh \left[ \frac{1}{\varphi_0} \lg \left( \frac{f_n}{f_0} \right) \right] + 1 \right\}, \quad (78)$$

with  $\varphi_0 = 0.2$  and  $f_0$  the cutoff frequency in the low-noise mode.

A very thin top electrode of the sample adds to the signal due to the transmission of incident light producing heat by absorption in the sample or at the back electrode. A thermally coupled backside of the piezoelectric transducer results in higher heat losses and, thus, in a lower pyroelectric current at low frequencies. Differences in the thermal expansion coefficients of the layers introduce noise caused by piezoelectricity.

---

## Examples of Nondestructive Evaluation of Embedded Piezoelectrics

Integrated piezoelectric sensors and actuators are used for health monitoring of safety components and for condition monitoring. In mechanical engineering, numerous applications are conceivable for damping vibrations in order to increase machining quality and machine dynamics (Gaudenzi 2009). During fabrication, such sensor-actuator devices are subjected to numerous thermal and mechanical loads leading to partial depolarization and, thus, degradation of the piezoelectric properties. Therefore, a nondestructive technique for the evaluation of polarization

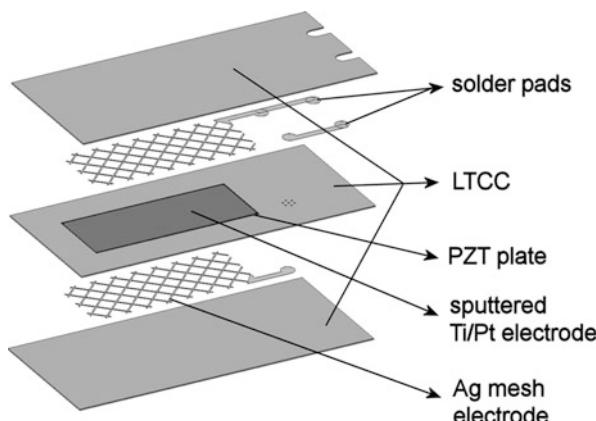
degradation in embedded piezoelectrics is an important issue for mass fabrication of such devices. A fundamental role in fabrication of such integrated piezoceramic sensors and actuators is also dedicated to performance inspection during the processing of laminated modules.

Recently, it was demonstrated that periodic heating by a modulated laser beam or laser pulses provides a promising approach for nondestructive evaluation of the polarization state in integrated sensor-actuator modules, comprising (i) piezoelectric PZT slabs embedded in low-temperature co-fired ceramics (LTCC) (Suchaneck et al. 2011, 2013a; Eydam et al. 2014, 2016a); (ii) macro-fiber composite (MFC) actuators consisting of PZT rods embedded in epoxy resin, placed between two copper electrodes and covered by a Kapton® film (Suchaneck et al. 2012; Eydam et al. 2014, 2016a); (iii) thermoplastic-compatible piezoceramic modules (TPM) (Eydam et al. 2014, 2016c); and (iv) PZT slabs coated with polyimide and embedded in casted aluminum (Suchaneck et al. 2013a; Eydam et al. 2015).

## LTCC/PZT Sensor-Actuator Modules

The manufacturing technology of novel LTCC/PZT modules was described in detail in Flössel et al. (2010) and Suchaneck et al. (2011). Figure 5 shows the schematic design of the LTCC/PZT module. The packaging technology uses the lamination of sintered and electroded PZT ceramic plates with LTCC green layers and subsequent sintering. Because of its anisotropic shrinkage behavior, Heraeus HeraLock® Tape-HL2000 (HL2000) turned out to be the most suitable LTCC material (Sea Ceramics Technology 2017b). As illustrated in Fig. 5, LTCC/PZT modules comprise a triple LTCC laminate setup. A cavity of the size of the PZT plate was cut into the middle LTCC green layer by laser machining. There the PZT plate was integrated. Green thickness of the middle LTCC layer had to be adjusted to reach a sintered thickness equal to the one of the PZT plate. Ag mesh electrodes were applied to the top and

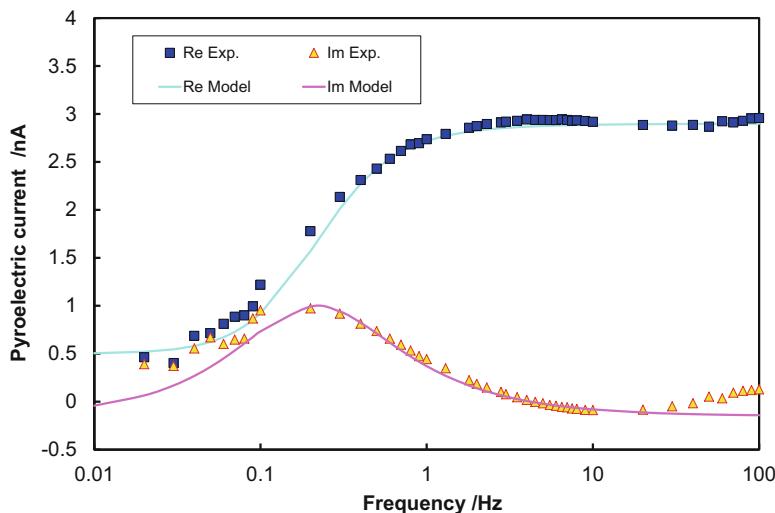
**Fig. 5** Schematic layout of the piezoelectric LTCC/PZT module (Suchaneck et al. 2011). (Reprinted with permission of Taylor & Francis)



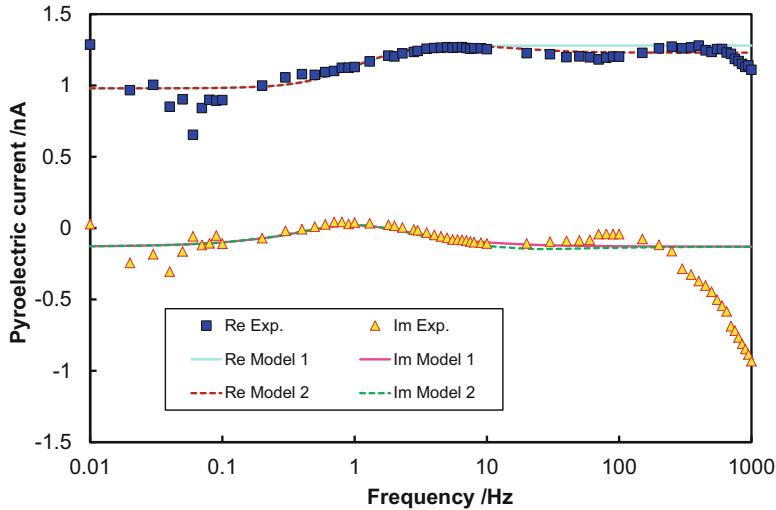
bottom LTCC green layers by screen printing. Vias filled with silver were formed for internal electrical connection to the two outer solder pads. The sample capacitance amounted to 30 nF and the dielectric loss tangent to about 2% at 10 kHz.

For comparison, a non-embedded PZT plate with a size of  $10 \times 7 \times 0.2$  mm<sup>3</sup> covered by a nickel-chromium electrode of about 100 nm thickness with a capacitance of 6.4 nF and a loss tangent of 1.8% at 10 kHz was analyzed. Figure 6 shows the pyroelectric current spectrum of the non-embedded PZT plate for 30 measurement repetitions which was fitted to Eq. (63). The thermal relaxation time, Eq. (24), amounts to 0.8 s corresponding to an interface thermal conductance of  $2G = 68$  W/m<sup>2</sup> K<sup>-1</sup>. This compares with a value of 50 W/m<sup>2</sup> K<sup>-1</sup> obtained by data fitting at the surface of electroded PVDF films (Bauer and Ploss 1988). The fit manifests a deviation from an ideal Debye-like model due to a continuous relaxation time distribution which is attributed to the impact of the electrodes and the electrical contact by a wire. The spectrum is best described by a Cole-Davidson relaxation with  $b = 0.9$ . The PZT plate shows a homogeneous polarization distribution. The obtained pyroelectric spectrum is comparable to the results in Bauer and Ploss (1990). Note that this analytical model was originally developed for non-embedded samples. However, the following demonstrates that the same model can be successfully applied also to embedded piezoelectric transducers.

Embedding the piezoelectric slab into the LTCC packaging, one has to take into account the strong absorption of laser radiation in the top LTCC layer (the absorption length of LTCC HL2000 in the wavelength region of 900 to 1000 nm is about 22  $\mu\text{m}$ ), the attenuation and phase shift of the temperature oscillation in the top layer, and thermal conduction loss from the front and back faces of the PZT slab (Mandelis



**Fig. 6** Pyroelectric current spectrum of a PZT plate in comparison to a fit to Eq. (63). Re and Im denote real and imaginary parts of the pyroelectric current, respectively (Eydam et al. 2016a)



**Fig. 7** Pyroelectric current spectrum of the piezoelectric LTCC/PZT module in comparison with fits to Eqs. (60) for model 1 and (61) for model 2, respectively. Re and Im denote real and imaginary parts of the pyroelectric current, respectively. (Adapted from (Eydam et al. 2016a))

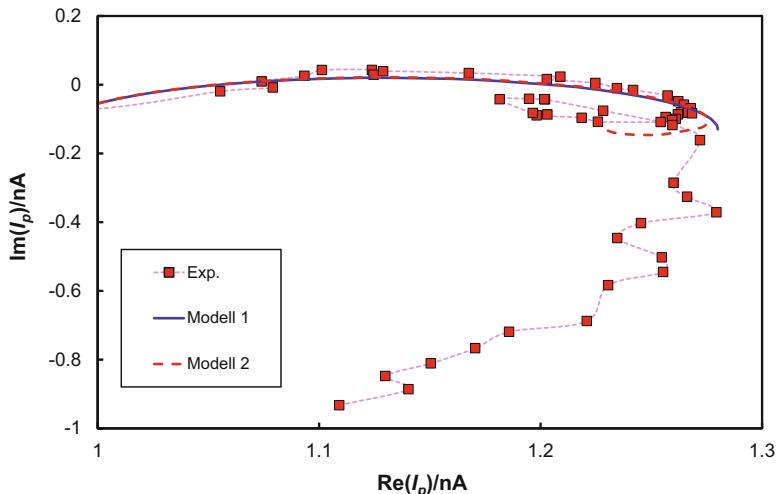
and Zver 1985). Figure 7 illustrates the pyroelectric current spectrum of the piezoelectric LTCC/PZT module in dependence on frequency for 30 measurement repetitions in comparison with fits to Eqs. (60) for model 1 and (61) for model 2, correspondingly. A minor decrease of the real part at higher frequencies was attributed to a slightly inhomogeneous polarization distribution (model 2). Following Eq. (65), one obtains  $p_1 = -p_0/5$ . Here, the data below 0.1 Hz do not fulfill the Rose criterion (cf. section “[Spatial Resolution](#)”). The signal-to-noise ratio increases continuously reaching about 65 at a modulation frequency of 10 Hz. Following Eq. (75), the spatial resolution becomes  $\delta z \approx 0.34z$ .

Above 6 Hz, the temperature oscillations are confined within the sample (cf. the half-wavelength criterion in section “[Temperature Oscillations as Diffusion Waves](#)”), and at frequencies above 25 Hz, only the top LTCC layer is heated. At low modulation frequencies  $f$  below 5 Hz, the pyroelectric response of the embedded PZT is governed by thermal losses induced by heat conduction to the embedding layers. Here, the sample behavior can be described by a harmonically heated piezoelectric plate exhibiting heat losses to the environment characterized by discrete relaxation times or even by their continuous distribution. The further increase of the pyroelectric current beyond 5 Hz up to a frequency of 10 Hz characterizes the decrease of thermal losses of the PZT plate when thermal loss cannot follow the increasing periodic heating rate. Beyond a frequency of 10 Hz, the pyroelectric response is described by a pyroelectric capacitor periodically heated by temperature oscillations from the LTCC top layer absorbed in the PZT plate top electrode. Above 100 Hz, the penetration depth of the temperature oscillations into LTCC becomes less than the laser light absorption depth determined by optical

transmission measurements on very thin LTCC samples. Consequently, the simplified model of harmonic heating from a very thin surface layer is invalid. The recorded pyroelectric current takes an additional time lag, i.e., a property of inertia, since time is needed for the flow of thermal energy from the LTCC top layer into the piezoelectric material. A similar behavior was obtained for heat flow in fluid systems exhibiting convection currents (Bosworth 1946). In electrical engineering, thermal inertia corresponds to inductive behavior.

Figure 8 depicts the Nyquist plots of the pyroelectric response of the investigated LTCC/PZT module. The maximum of the Nyquist plot curves satisfies the condition  $\omega\tau = 1$ . This corresponds to a thermal relaxation time  $\tau_{th}$ , Eq. (24), of 0.16 s resulting assuming  $c = 2.7 \text{ MJ/m}^3 \text{ K}$  (Suchanek et al. 2012; PI Ceramic GmbH 2017) in a total thermal conductance of  $2G = 337.5 \text{ W/m}^2 \text{ K}^{-1}$ , i.e., about  $170 \text{ W/m}^2 \text{ K}^{-1}$  at both the top and the bottom surface. This compares with a value of  $160 \text{ W/m}^2 \text{ K}^{-1}$  obtained for screen-printed LTCC directly deposited onto an aluminum substrate (Mah et al. 2017). The long tail with  $\Im(I_p) < 0$  represents the inductive behavior of the thermal system described above.

In 1983, a technique called transmission thermal wave microscopy was introduced to detect and image hidden subsurface thermal inhomogeneities of materials (Baumann et al. 1983). The thermal image is formed by scanning a point heat source (an electron beam) over the surface of the sample. The temperature oscillation propagating through the sample is directly detected by a pyroelectric sensor in intimate thermal contact with the sample. Propagating directly through a solid sample, it traverses a shorter distance to the sensor than a temperature oscillation interacting with obstacles. This forms a pyroelectric contrast. Our case is very similar



**Fig. 8** Nyquist plot of the pyroelectric response of the piezoelectric LTCC/PZT module. Squares, experiment; solid line, fit to a homogeneous polarization, Eq. (60); and dashed line, considering a slightly non-homogeneous polarization following Eq. (61). Re and Im denote real and imaginary parts of the pyroelectric current, respectively

but uses a laser as a heat source. Here, thermal relaxation characterizes the thermal contact and, thus, enables locating lamination failures. For calculation, an ideal thermal contact, Eq. (30), was assumed. Thus, in the presence of an air gap (cf. Fig. 9), the thermal interface conductance  $G$  decreases and, consequently, the thermal relaxation time  $\tau_{th}$ , Eq. (24) increases. Using a 2D scanning system, one arrives to the Scanning Laser Intensity Modulation Method described above in section “Two-Dimensional and Three-Dimensional LIMM.”

## Macro-Fiber Composite Actuators

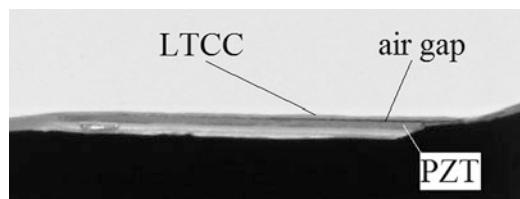
Macro-Fiber composites (MFC) were developed at NASA Langley Research Center (Wilkie et al. 2000). They use both active and composite material technologies in addition to electrodes. Rectangular piezoceramic rods are sandwiched between layers of adhesive, electrodes, and polyimide film. An electrode pattern transfers the applied voltage directly to and from the ribbon-shaped rods (Fig. 10).

The particular sample considered here is a commercial M-8528-P2 piezoelectric actuator (Smart Materials GmbH, Dresden, [www.smart-materials.com](http://www.smart-materials.com)) with an overall length of 105 mm, an active length of 85 mm, an active width of 28 mm, and a thickness of about 0.3 mm. The sample capacitance was 170 nF with a dielectric loss tangent of about 5% at 10 kHz.

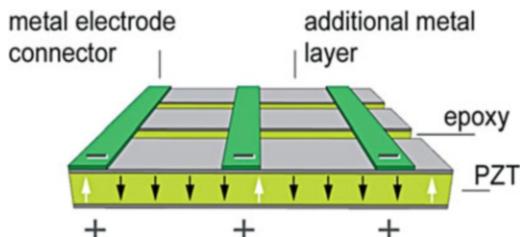
By applying a voltage, the MFC bends, counteracting or generating vibrations. Without an applied voltage, the MFC represents a very sensitive strain gauge, sensing deformations, noise, and vibrations or is used as a vibration energy harvester.

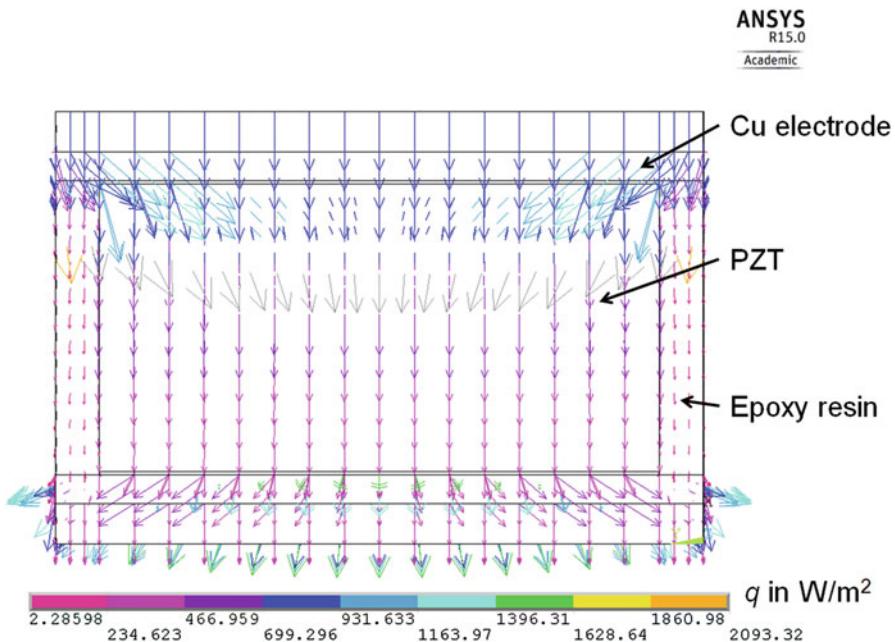
Although MFCs exhibit a very complex structure, numerical simulations revealed a significant blurring of the temperature field by the metallic electrodes and a small horizontal component of the temperature gradient at the piezoceramics/epoxy resin

**Fig. 9** Air gap between the PZT plate and the LTCC layers



**Fig. 10** Schematic structure of a piezoelectric macro-fiber composite actuator. (© Smart Material Corp.)



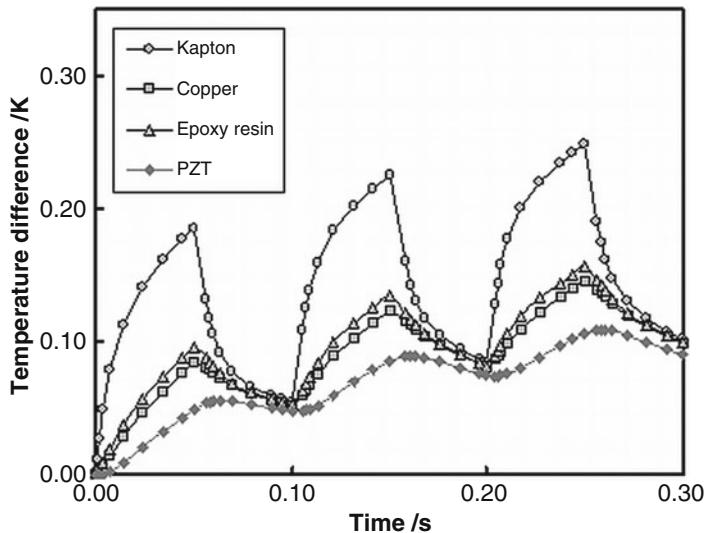


**Fig. 11** Heat flux distribution in the MFC transducer at  $t = 0.15$  s for a heat flux modulation frequency of 10 Hz

interface (Fig. 11), that is, heat conduction through this interface can be neglected (Suchaneck et al. 2012). Consequently, expansive finite element modeling of this complex thermal problem may be avoided.

The embedded piezoelectric rods are subjected to periodic heating by a square wave-modulated laser array through the top layers. Transient thermal analysis revealed the presence of a transient heating-up period before reaching the steady state (Fig. 12) (Suchaneck et al. 2012). Inside the PZT piezoceramics, the thermal excitation becomes nearly sinusoidal and shifted in phase as known for harmonic excitation. At low modulation frequencies, the pyroelectric response of the embedded PZT rods is governed by thermal losses to the embedding layers. Therefore, the sample behavior can be described by harmonically heated piezoelectric rods exhibiting heat losses to the environment that can be characterized by a single thermal relaxation time, Eqs. (24 and 60). This enables the estimation of the thermal conductance at the interfaces of the embedded piezoelectric.

The pyroelectric response of the MFC M-8528-P2 actuator is satisfactorily modeled by assuming a homogeneous polarization. It is well described by a Cole-Cole relaxation, Eq. (63), with  $a = 0.4$ ,  $b = 1$ , with a broad distribution of relaxation times around  $\tau_{th} = 5.3$  s, and a single relaxation time at  $\tau_{th} = 1.6$  ms (Figs. 13 and 14). According to Eq. (24), the first value describes the heat loss of the sample surface to the environment with a thermal conductance of about



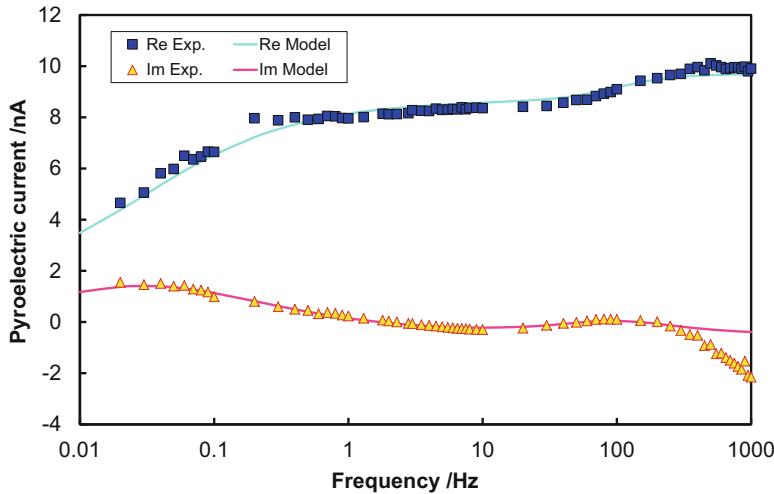
**Fig. 12** Time evolution of the temperature difference to the environment in the center of the different MFC actuator materials for a heat flux modulation frequency of 10 Hz (Suchanek et al. 2012). (Reprinted with permission of IEEE)

100 W/m<sup>2</sup> K<sup>-1</sup>. The smaller relaxation time corresponds to an interface thermal conductance of approximately 3·10<sup>4</sup> W/m<sup>2</sup> K<sup>-1</sup> which is characteristic for interfaces between two thin metal sheets (Chen et al. 2014). However, this approximate value of  $G$  is of large uncertainty since at frequencies above 250 Hz, only a surface layer of less than 70 µm is heated and the simple model of harmonic surface heating is no more valid. Here, heat conductivity spreads the heat energy deposited into the instantaneously heated surface layer throughout the remaining volume resulting in an inductive thermal behavior. Similar behavior was obtained for the LTCC/PZT module (cf. Fig. 8). For the MFC M-8528-P2 actuator, all data satisfied the Rose criterion (cf. section “[Spatial Resolution](#)”). Above a modulation frequency of 2 Hz, the spatial resolution defined by Eq. (75) was  $\delta z \approx 0.27z$ .

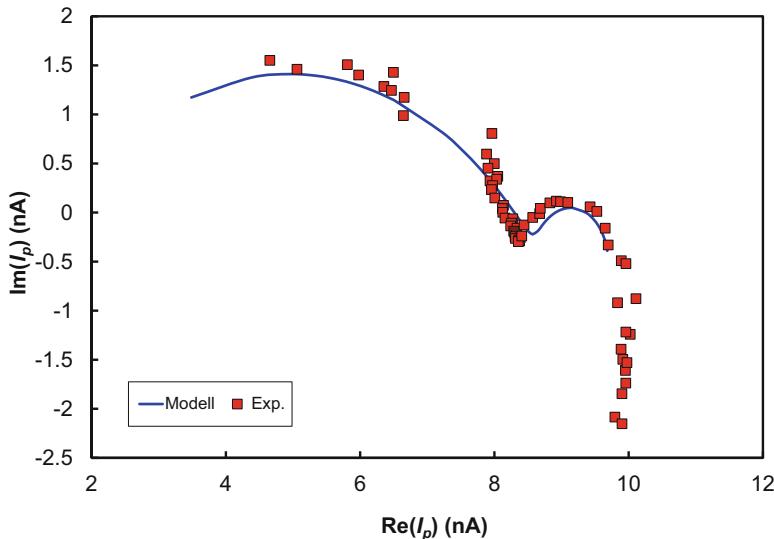
### Thermoplastic-Compatible Piezoceramic Modules (TPM)

The manufacturing process of TPMs is illustrated in Fig. 15 (Hufenbach et al. 2013).

The base materials of the process are two thermoplastic film webs. TPM fabrication starts with the deposition of the electrodes onto the two thermoplastic films, e.g., by screen printing. The piezoelectric patch is then positioned on the electrode on the lower web and covered with an upper thermoplastic film. The sample considered here exemplarily consisted of a PZT plate with a size of 35 × 22 × 0.2 mm<sup>3</sup>. It was positioned between two transparent thermoplastic carrier

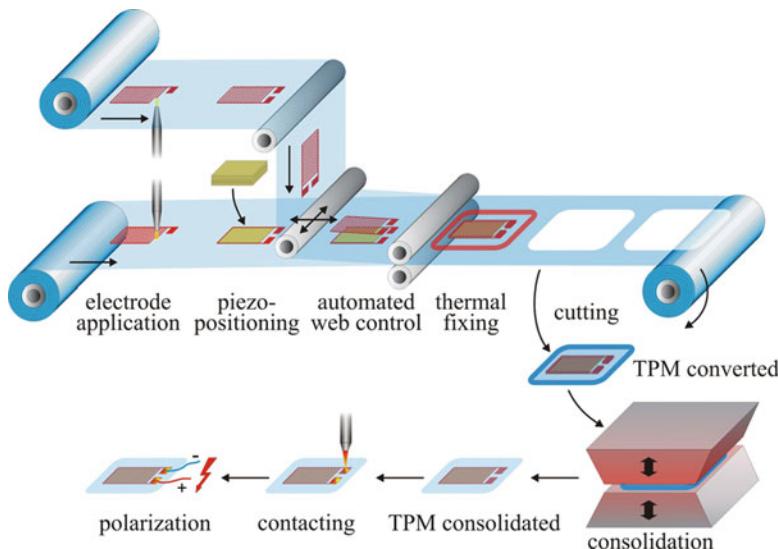


**Fig. 13** Pyroelectric current spectrum of the piezoelectric MFC M-8528-P2 actuator in comparison with a fit to a Cole-Cole relaxation and a very short single relaxation time. Re and Im denote real and imaginary parts of the pyroelectric current, respectively (Eydam et al. 2016a)



**Fig. 14** Nyquist plot of the pyroelectric response of the MFC M-8528-P2 actuator in comparison with a fit to a Cole-Cole relaxation and a very short single relaxation time. Re and Im denote real and imaginary parts of the pyroelectric current, respectively

films of polycaprolactam (PA6) with a thickness of 100  $\mu\text{m}$  covered by a grid electrode. The capacitance of the sample was 70 nF, the dielectric loss tangent amounted to about 2% at 10 kHz. The aligned components are fixed by a thermal



**Fig. 15** Schematic diagram of the TPM manufacturing process (Hufenbach et al. 2013)

welding step and punched out. After a consolidation step, contacts are provided and the piezoelectric patch is poled. Figure 16 shows a TPM in its final state.

Also here, all measurement data satisfied the Rose criterion (cf. section “[Spatial Resolution](#)”). Above a modulation frequency of 10 Hz, the spatial resolution defined by Eq. (75) was  $\delta z \approx 0.15z$ .

Figure 17 presents the pyroelectric current spectrum of the TPM in comparison with a fit to a superposition of several relaxation times (Table 1).

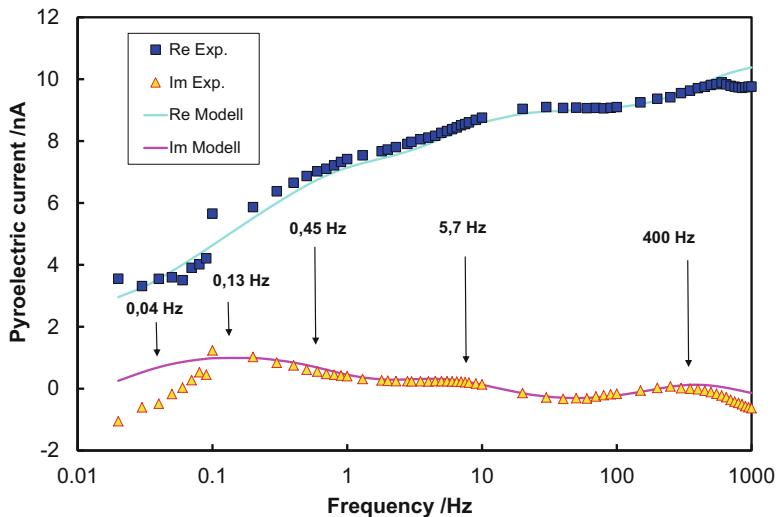
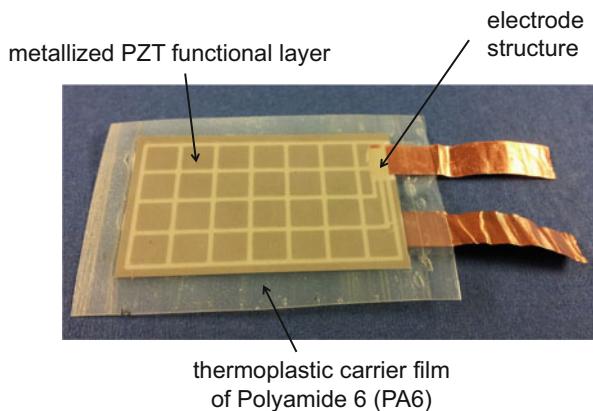
The scattering of the measured signal below 0.1 Hz is attributed to a higher measurement uncertainty at low frequencies. Therefore, the first relaxation time is very uncertain, but it is needed to achieve a good fit to experimental data.

The second relaxation time  $\tau_2$  is caused by thermal losses of the sample top and bottom to the environment. Here, the interface thermal conductance of  $2G = 130 \text{ W/m}^2 \text{ K}^{-1}$  (cf. Eq. (24)) is comparable to a value of  $50 \text{ W/m}^2 \text{ K}^{-1}$  obtained by data fitting at the surface of electroded PVDF films (Bauer and Ploss 1988).

The third relaxation time  $\tau_3$  was attributed to the thermal resistance of the PZT/PA6 interface. A nearly identical time constant was found in a similar configuration of a metalized pyroelectric triglycine sulfate slab with a thickness of 0.1 mm placed on a copper block (Hadni et al. 1969).

The fourth relaxation time corresponds to the bulk thermal resistance  $R_{th,b} = \kappa/d$  through the PA6 film. The shortest relaxation time we related to the metal-metal interface between the metalized PZT and the screen printed electrode structure with respect to a similar value in the MFC transducer (cf. section “[Macro-Fiber Composite Actuators](#)”).

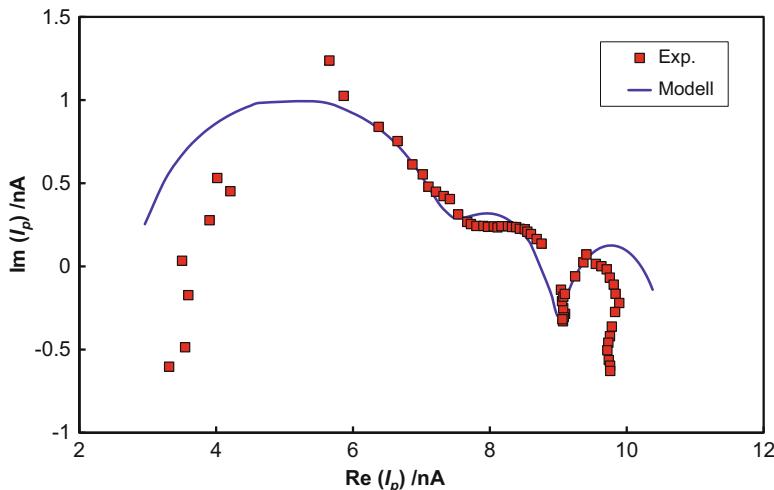
**Fig. 16** A thermoplastic-compatible piezoceramic module in its assembled state



**Fig. 17** Pyroelectric current spectrum of the TPM in comparison with a fit to a superposition of five relaxation times each described by Eqs. (24 and 60). Re and Im denote real and imaginary parts of the pyroelectric current, respectively

**Table 1** Thermal relaxation times and origin of thermal losses

Time constant	Relaxation time, ms	Origin of thermal loss
$\tau_1$	~4000	Unknown
$\tau_2$	1500	Sample top/ environment
$\tau_3$	350	PZT/PA6 interface
$\tau_4$	30	Bulk thermal conductance of PA6
$\tau_5$	0.7	Pt/Ag-paint interface



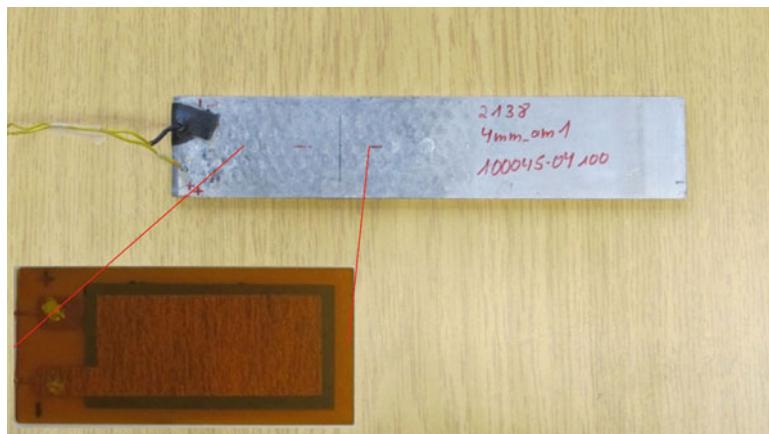
**Fig. 18** Nyquist plot of the pyroelectric response of the TPM. Squares, experiment; solid line, fit as a superposition of five relaxation times each described by Eqs. (24 and 60). Re and Im denote real and imaginary parts of the pyroelectric current, respectively

Figure 18 illustrates the corresponding Nyquist plot. As for the other samples, an inductive behavior occurs also for the TPM sample at high laser modulation frequencies where only the sample surface is heated. This gives evidence of thermal energy accumulation before the heat flux reaches the pyroelectric transducer.

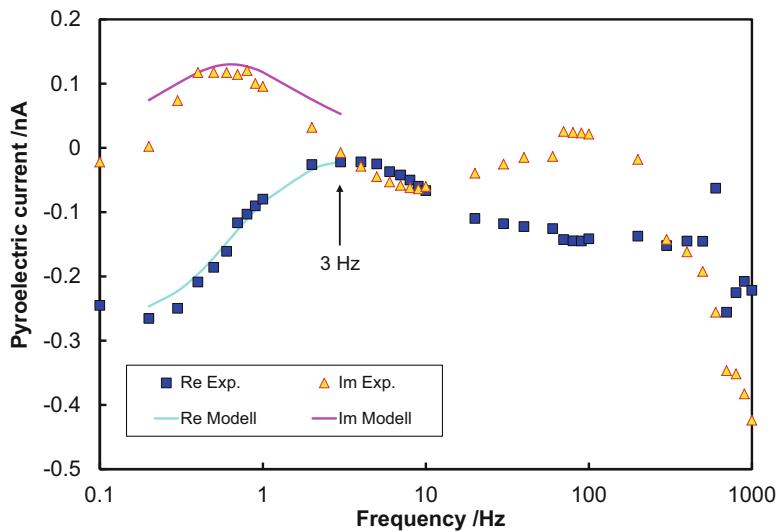
### Integrated Cast Al/PI/PZT Piezomodules

Integrated piezoceramic sensor-actuators were manufactured by high-pressure die casting. DuraAct<sup>TM</sup> piezoelectric patch transducers P876K001 of  $39.5 \times 19 \times 0.5 \text{ mm}^3$  size consisting of a  $15 \times 30 \times 0.2 \text{ mm}^3$ -sized piezoceramic plate and a polyimide package (PI Ceramic GmbH, Lederhose, Germany) were fixed between sheets of expanded metal inside the mold onto two ejector pins. Casting was performed employing the casting alloy AlSi9Cu3Fe ( $c = 2.6 \text{ MJ/m}^3\text{K}$ ,  $\kappa = 115 \text{ W/mK}$  (Raffmetal (2017))) at a mold temperature of  $150^\circ\text{C}$  and a maximum plunger velocity of  $2 \text{ m/s}$  during filling. The casted plate had a dimension of  $178 \times 178 \times 4 \text{ mm}^3$ . The piezoelectric patch was located in the center at about  $20 \text{ mm}$  from the upper edge. Samples with a width of about  $20 \text{ mm}$  were cut out from the casted plate (Fig. 19).

Details of the fabrication process were reported in Rübner et al. (2008). The piezoelectric patch capacitance was  $25 \text{ nF}$ ; the dielectric loss tangent amounted to about  $3\%$  at  $10 \text{ kHz}$ . Samples with patch transducers located at the neutral axis and displaced off-axis, respectively, were prepared. Figure 20 shows the pyroelectric



**Fig. 19** Integrated die-casted Al/PI/PZT piezomodule



**Fig. 20** Pyroelectric current spectrum of a piezoelectric die-casted Al/polyimide/PZT module. Re and Im denote real and imaginary parts of the pyroelectric current, respectively (Eydam et al. 2015). (Reprinted with permission of Taylor & Francis)

current spectrum of the die-casted Al/polyimide/PZT piezoelectric module obtained by 50 measurement repetitions for averaging.

The value of the pyroelectric current is in the range of a few 100 pA owing to the large heat capacity of the embedding cast aluminum. Up to 3 Hz, the SNR of 3 to 9 is around the Rose limit. Due to the larger heat capacity of the cast aluminum embedding, thermal losses occur already at one order of magnitude lower frequencies. According to the half-wavelength criterion in section “Temperature Oscillations

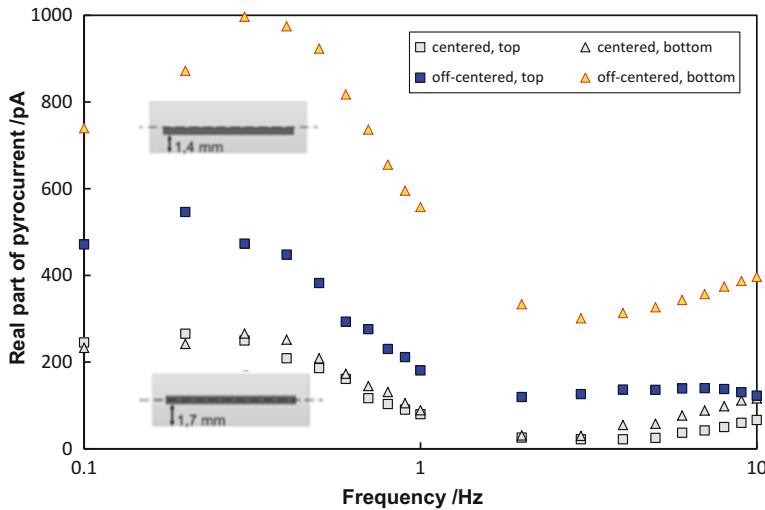
as Diffusion Waves,” the temperature oscillation reaches the back side of the pyroelectric patch at about 1 Hz and the front side at 3 Hz. For frequencies above 9 Hz, solely the top aluminum plate is heated. Above 200 Hz, the pyroelectric current is shifted in phase additionally up to  $\pi/4$  due to a surface layer of a 300 to 400 mm thickness exhibiting different thermal properties (Eydam et al. 2014).

The first peak in the imaginary part of the pyroelectric current reveals a thermal relaxation time of  $\tau = 0.2$  s describing heat losses of the PZT plate to the embedding polyimide. The thermal conductance at the interface amounts to  $2G = 2730 \text{ W/m}^2 \text{ K}^{-1}$ , calculated using Eq. (24). Since both polyimide and polyamide are thermally comparable polymers, the value of  $G$  is justified by a similar configuration of a metalized triglycine sulfate plate with a thickness of 0.1 mm on a 4-mm-thick mylar sheet possessing  $G = 1800 \text{ W/m}^2 \text{ K}^{-1}$  (Hadni and Thomas 1972).

According to Eq. (5), the propagation velocity of thermal oscillations is about 25 mm/s at a laser modulation frequency of 1 Hz. Consequently, the piezoelectric slab is heated at low frequencies from both sides with very small time lag. Here, the signs of the real and imaginary parts of the pyroelectric currents are sensitive to an off-axis location of the piezoelectric patch. The off-center position of the piezoelectric device leads to differences in amplitude and phase (not shown here) between the measurements from each side. This can be used to detect off-centering of the piezoelectric patch. To avoid ambiguity, the following considers the absolute value of the real part of the pyroelectric current spectrum for off-center detection.

Two samples were compared: a center-positioned transducer (c) and a transducer displaced off-center (oc). Both samples were irradiated both from the top side and the bottom side. The results are illustrated in Fig. 21. In the case of an ideally center-positioned transducer, the measurements from both sides agree well with each other. An off-centered position of the piezoelectric transducer leads to noticeable differences between these two measurements. For a smaller depth of transducer location, a larger pyroelectric current was obtained. Moreover, the different positions of the maxima of the pyroelectric current’s real part characterize the different thermal wave penetration depth up to the transducer’s center of polarization. Hence, it is possible to verify the centricity or the eccentricity of the embedded piezoceramic slabs by measuring the pyroelectric response from both the top and the bottom side. Similarly, charge centroids were determined by means of the thermal pulse method in electron beam-charged PTFE films with a thickness of 25  $\mu\text{m}$  (von Seggern et al. 1984).

As for the other samples, the Nyquist plot is an appropriate tool to analyze the origin of the frequency response (Fig. 22). The data is well described by a Debye relaxation up to a laser modulation frequency of 3 Hz. Since at frequencies above 3 Hz only the aluminum cover layer is heated, an inductive behavior occurs similarly to the LTCC/PZT and MFC samples. This is consistent with a phase jump of about  $\pi/2$  in the region of 1 to 4 Hz (Eydam et al. 2015).



**Fig. 21** Absolute value of the real part of the pyroelectric current spectrum of an Al/PI/PZT module with centered and off-centered transducer location and for irradiation from the top and bottom sides, respectively (Eydam et al. 2015). (Reprinted with permission of Taylor & Francis)

## Examples of Thermal Pulse Measurements

### Evaluation of Embedded Piezoelectric Sensor-Actuators

When thermal pulses are applied with a pulsed laser, the signal  $U(t_l)$  is recorded in the time domain and converted to the frequency domain by a discrete fast Fourier transform (Mellinger et al. 2005a). The obtained frequency spectrum is divided by the transfer function  $H(f_n)$  of the measurement setup to account for the influence of the amplifier settings. The resulting pyroelectric current spectrum yields:

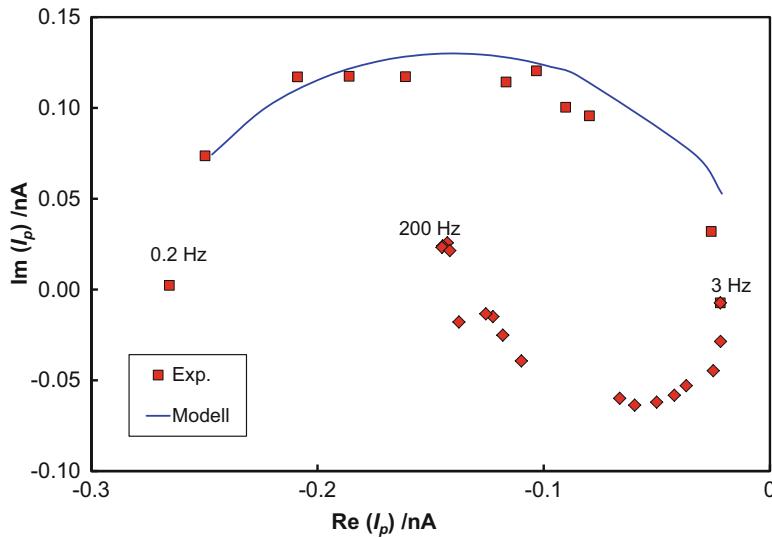
$$I(f_n) = \frac{\Delta t}{H(f_n)} \sum_{l=0}^{N-1} U(t_l) \exp\left(-inl\frac{2\pi}{N}\right), \quad (79)$$

with

$$f_n = \frac{n}{N\Delta t}, \quad n = 0, \dots, N-1, \quad (80)$$

where  $l$  is the index of the sampled signal values,  $n$  is the index of the Fourier coefficients,  $N$  is the number of data points (both in time domain and in frequency domain), and  $\Delta t$  is the sampling rate.

Thermal pulse measurements were performed on the samples described above using the four different SR570 preamplifier settings of Table 2 to cover a wide



**Fig. 22** Nyquist plot of a piezoelectric die-casted Al/polyimide/PZT module. Re and Im denote real and imaginary parts of the pyroelectric current, respectively (Eydam et al. 2015). (Reprinted with permission of Taylor & Francis)

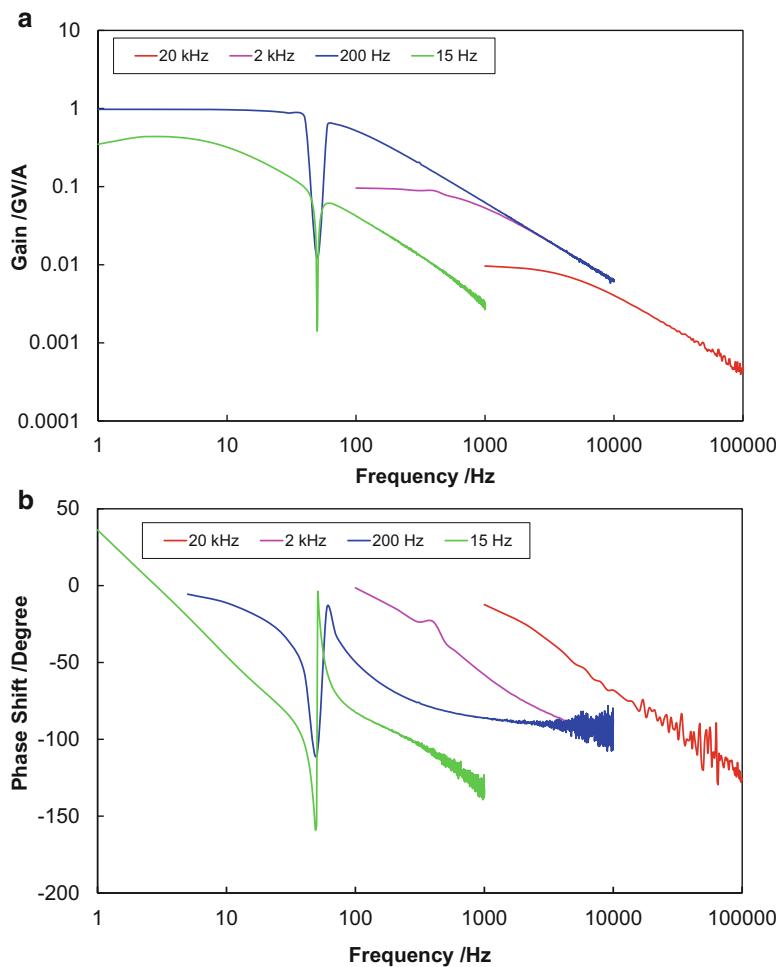
**Table 2** Preamplifier settings and sampling rates for thermal pulse measurements

Number of measurement	Gain mode	Gain, V/A	Bandwidth, Hz	Sampling rate, s
1	High-bandwidth	$10^7$	20,000	$10^{-8}$
2	High-bandwidth	$10^8$	2000	$10^{-7}$
3	High-bandwidth	$10^9$	20	$10^{-6}$
4	Low-noise	$10^9$	15	$10^{-5}$

bandwidth range at low noise. A pulse width of 100 ns was chosen. The number of sampled signal values amounted to approximately  $N = 10^5$  for each measurement.

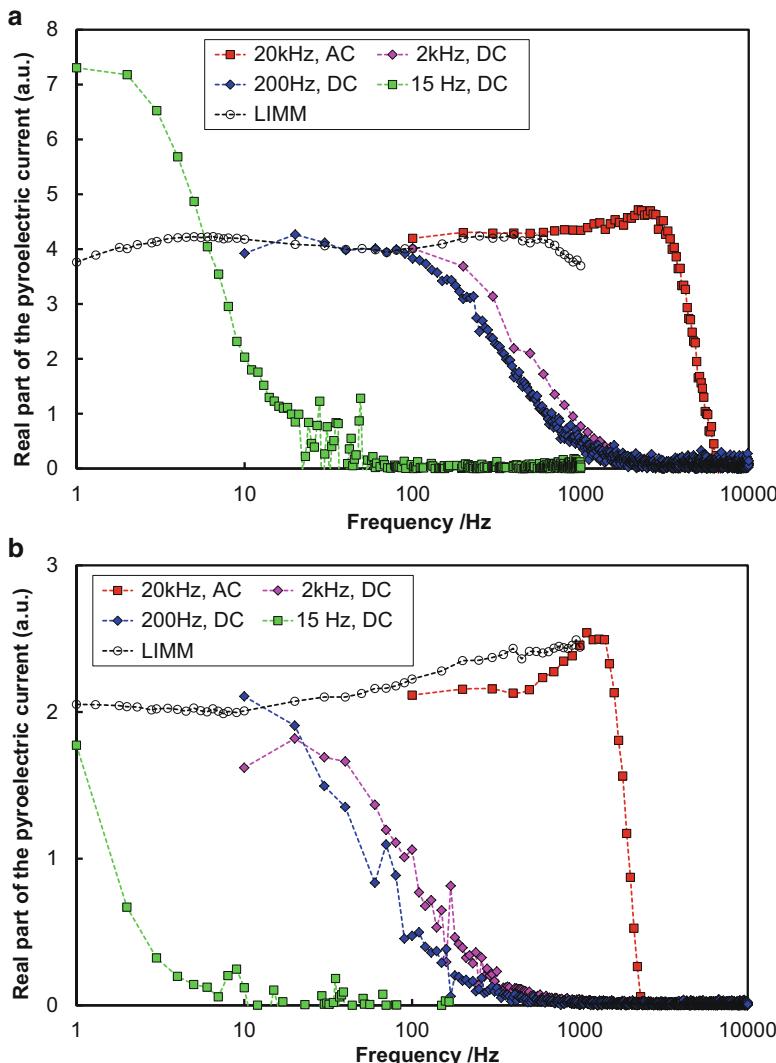
The resulting spectra were divided by the corresponding transfer functions of the measuring setup. The transfer functions of the different preamplifier settings and an additional 50 Hz notch filter were determined in advance. They are depicted in Fig. 23. Above the cut-off frequency of the preamplifier, the gain decreases by about 20 dB per decade. The filter induces a notch in the amplitude response and a phase jump at 50 Hz. Note that for 15 Hz, a high-pass filter with a cut-off frequency of 1 Hz was added to reduce DC-offsets. For better comparison, the obtained curves (Fig. 24) were normalized by multiplying with  $\Delta t$ . Since the imaginary and real parts of consistent pyroelectric spectra are interrelated by the Kramers-Kronig relations (Tuncer and Lang 2006), only the real part was considered.

Figure 24 illustrates that the decreasing gain and the phase shifts of the chosen current amplifier at frequencies above the cut-off frequency lead to a signal distortion at high frequencies regardless of the corrections made. This is attributed to large capacitive loads not accounted for by the measurement of the transfer function. The capacitive load of the PZT plate of about 7 nF is acceptable. In this case, there are only minor changes in the transfer function. Capacitive loads above 100 nF characteristic for the embedded transducers produce large distortions. Here, the transfer function including the capacitive load cannot be determined experimentally since the load is a part of a capacitive voltage divider causing here an overload of the preamplifier current. Therefore, the LIMM spectrum in a broad frequency region will

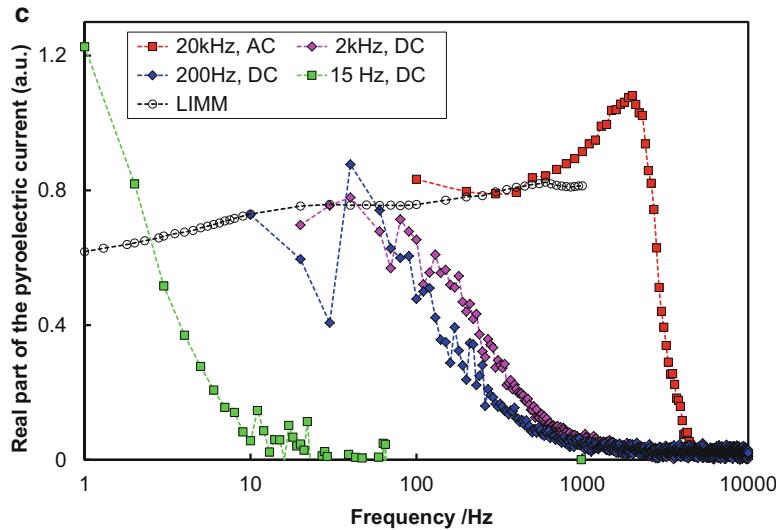


**Fig. 23** Amplitude (a) and phase transfer functions (b) of the amplifier SR570 including the 50 Hz notch filter without an external load for the settings in Table 2. (Adapted from (Eydam et al. 2016c))

be reproduced only by several measurements, each optimized for a gain-bandwidth relation. Thus, the thermal pulse method is beneficial if information is required in a limited frequency range, e.g., in the frequency range around 10 Hz where the pyroelectric response is determined by the polarization of the embedded piezoelectric plate. By this way, degradation of the piezoelectric properties during the integration process caused by thermal and mechanical loads is easily and quickly detected.



**Fig. 24** (continued)

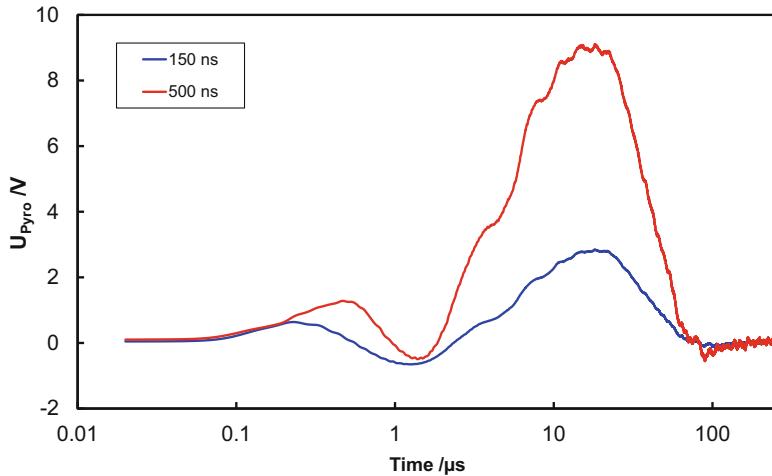


**Fig. 24** Real parts of the pyroelectric current spectra for an LTCC/PZT module (a), the MFC M-8528-P2 actuator (b), and a thermoplastic-compatible piezoceramic module (c), all determined by a fast Fourier transform of the thermal pulse response in comparison with the LIMM spectrum

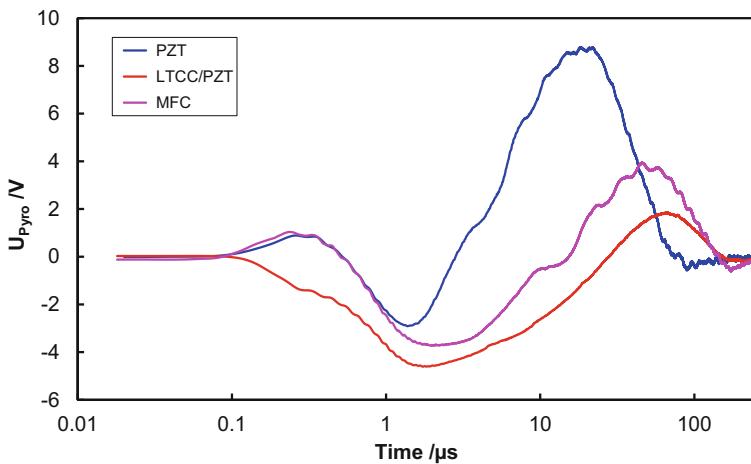
### Additional Information Derived from the Thermal Pulse Method at Very Short Response Times

For comparison purposes, a non-embedded PZT plate with a size of  $10 \times 7 \times 0.2 \text{ mm}^3$  covered by a nickel-chromium electrode (cf. section “[LTCC/PZT Sensor-Actuator Modules](#)”) was analyzed by a pulse train with a repetition frequency of 1 Hz. Figure 25 illustrates that initially the heating rate of the piezoelectric plate, which is proportional to the pyroelectric signal (cf. Eq. (12)), increases during pulse duration. At the beginning of the pulse-off time, the heating rate decreases even with cooling for a very short period of time followed by a large increase where a longer pulse width leads to a higher heat input and, thus, a larger signal (Eydam et al. 2016a). Here, the rising signal is distorted by the limited bandwidth of the measuring setup. The maximum value of the pyroelectric signal is located approximately at the inverse of the bandwidth. It is followed by an almost linear decay up to the point of changing sign. This behavior is similar to ultrathin (10 nm) PVDF films irradiated with 10-ps-duration pulses which was verified by numerical simulations of the thermal problem (Coufal and Grygier 1989).

The decrease of the heating rate immediately after the pulse active time is attributed to the thermal buffering effect of the top electrode (Bloß et al. 2000). It absorbs thermal energy during the short thermal pulse but only slowly transfers it to the piezoelectric plate since the heat transfer is limited by the propagation speed of



**Fig. 25** Time dependence of the output voltage of the current amplifier ( $10^8$  V/A, 7 kHz bandwidth) for the PZT plate for an optical power of 22.5 W, a repetition frequency of 1 Hz, and pulse widths of 150 ns and 500 ns, respectively (Eydam et al. 2016a)



**Fig. 26** Time dependence of the output voltage of the current amplifier ( $10^8$  V/A, 7 kHz bandwidth) for the PZT plate, the LTCC/PZT sensor-actuator, and the MFC actuator for an optical power of 65 W, a repetition frequency of 1 Hz, and a pulse width of 150 ns (Eydam et al. 2016a)

acoustic phonons. The latter is related to thermal diffusivity which characterizes the ability of a material to conduct thermal energy. Note that the time domain shown in Fig. 25 is much lower than the thermal transit time of the PZT plate amounting to about 10 ms.

Figure 26 compares the pyroelectric response to the pulse train of the PZT plate, the LTCC/PZT sensor-actuator, and the MFC actuator. The attenuation of the

absorbed heat in the top layer results in a smaller maximum of the pyroelectric signal. However, quantitative estimates of the thermal properties of the top layer are not possible since the short-term signal is distorted by the bandwidth of the measurement setup. The short-term behavior of the MFC actuator follows that of the PZT plate since in both cases, the top electrode is directly heated by the laser radiation. For the MFC actuator, top electrode heating is accomplished through the nearly transparent polyimide film. In case of the LTCC/PZT actuator module, laser irradiation is absorbed in a LTCC surface layer of about 20  $\mu\text{m}$  thickness (cf. section “[LTCC/PZT Sensor-Actuator Modules](#)”).

---

## Conclusion

Thermal wave techniques can be used to excite thermal gradients within the volume to be evaluated and, hence, to achieve information on the interior of the sample under test. The opportunities of thermal waves have been considered for the particular case of embedded piezoelectrics and the evaluation of its polarization state in construction elements, e.g., adaptronic structures. The described methods for this application are based on the pyroelectric effect, whereas thermal excitation was induced by laser irradiation.

Piezoelectric sensors and actuators are provided with electrodes which transfer the voltage directly to and from the piezoelectric material. Therefore, they are easily accessible to electrical measurements. A smart characterization technique of piezoelectrics makes use of their pyroelectricity. For this purpose, the piezoelectric material is thermally excited and the resulting pyroelectric current is recorded. Measurements are feasible both in the frequency and the time domain.

The evaluation of the pyroelectric current spectrum is a simple and promising approach also for the nondestructive evaluation even of embedded piezoelectric transducers. The thermal pulse method yields similar results, but much faster. Additionally, the pyroelectric signal in the time domain provides information about heat transfer through the top layer and at the interfaces in the  $\mu\text{s}$  and  $\text{ms}$  range. Nevertheless, the realization of thermal pulse methods is restricted by gain/bandwidth tradeoffs of commercially available preamplifiers which should be adapted also to capacitive loads.

The spreading of the temperature oscillations and pulsed propagating through the sample rapidly degrade the resolution at larger depths. Thermal methods are beneficial in the near-surface region up to a depth of about 1 mm for an excitation frequency in the order of 1 Hz. At larger depths and very weak pyroelectric response, the pyroelectric current spectrum allows at least to distinguish the center and the off-center position of the transducer.

## Summary

Thermal wave techniques can be used to excite thermal gradients within the volume to be evaluated and, hence, to achieve information on the interior of the sample under test. If the surface of the sample under test is excited homogeneously then any deviations which influence the thermal conductance within this sample will lead to a non-homogeneous temperature field. This allows to detect and to reconstruct these inhomogeneities – e.g. thickness differences, delaminations, hidden holes – by related mathematical methods. Thermal wave excitation to generate these temperature fields is usually done by modulated thermal radiation, mostly by periodically or sinusoidally changing radiation or by thermal pulses.

This chapter focuses in particular on the determination of the polarization depth profile of piezoelectric modules by thermal methods. For that, the pyroelectric response of the piezoelectric elements is recorded both in the frequency and in the time domains. Spatial resolutions in the order of 100 nm are achievable by using high-frequency LIMM or thermal pulse techniques, whereby the highest resolution is achieved near the surface. It rapidly decreases toward the middle of the sample owing to the physics of thermal diffusion. The spatial resolution is lower for materials with a larger thermal diffusivity.

Thermal methods based on the pyroelectric effect are a promising approach for nondestructive evaluation of the polarization state in light-weight components comprising integrated piezoelectric sensors and actuators. Here, the analysis of the frequency spectrum by means of Nyquist complex plane mapping provides a more direct and easy understanding of the heat conduction processes. This approach is easily implemented into the robotic high-volume production of active structural components.

---

## References

- Adby PR (1980) Applied circuit theory: matrix and computer methods. Ellis Horwood series in electrical and electronic engineering. Ellis Horwood Ltd, London
- Agnel S, Toureille A, Le Gressus C (1996) Study of the aging of impregnated paper of high power capacitors using the thermal step method and the thermally stimulated currents. Proceedings of conference on electrical insulation and dielectric phenomena – CEIDP '96. <https://doi.org/10.1109/CEIDP.1996.564656>
- Ahmed NH, Srinivas NN (1997) Review of space charge measurements in dielectrics. IEEE Trans Dielectr Electr Insul 4:644–656
- Amjadi H (1996) Thermal-pulse investigation of thermally grown silicon dioxide electrets. In: 9th international symposium on electrets (ISE 9). <https://doi.org/10.1109/ISE.1996.578079>
- Angström AJ (1863) XVII. New method of determining the thermal conductivity of bodies. Philos Mag Ser 4 25(166):130–142
- Aryal S, Mellinger A (2013) Resolution-enhanced polarization imaging with focused thermal pulses. J Appl Phys 114:154109
- Bauer S (1993) Method for the analysis of thermal-pulse data. Phys Rev B 47:11049–11055
- Bauer S, Bauer-Gogonea S (2003) Current practice in space charge and polarization profile measurements using thermal techniques. IEEE Trans Dielectr Electr Insul 10:883–902

- Bauer S, Ploss B (1988) Analysis of the spatial distribution of polarization in PVDF-foils from the frequency spectra of the pyroelectric current. In: Proceedings of 6th international symposium on electrets (ISE 6). <https://doi.org/10.1109/ISE.1988.38519>
- Bauer S, Ploss B (1990) A method for the measurement of the thermal, dielectric, and pyroelectric properties of thin films and their applications for integrated heat sensors. *J Appl Phys* 68:6361–6367
- Baumann T, Dacol F, Melcher RL (1983) Transmission thermal-wave microscopy with pyroelectric detection. *Appl Phys Lett* 43:71–73
- Bezdety NM, Zeinaly AK, Khutorsky VE (1984) Investigation of the polarization distribution in ferroelectrics by a dynamic pyro-effect method [in Russian]. *Izv AN SSSR Ser Fiz* 48:200–203
- Biot MA (1970) Variational principles in heat transfer. Clarendon Press, Oxford
- Blevin WR, Geist J (1974) Influence of black coatings on pyroelectric detectors. *Appl Opt* 13:1171–1178
- Bloß P, DeReggi AS, Schäfer H (2000) Electric-field profile and thermal properties in substrate-supported dielectric films. *Phys Rev B* 62:8517–8530
- Bosworth RCL (1946) Thermal inductance. *Nature* 158:309
- Braccini M, Dupeux M (2012) Mechanics of solid interfaces. Wiley, Hoboken, chapter 8.3.1
- Burfoot JC, Latham RV (1963) A new method for studying movements of electric domain walls. *Br J Appl Phys* 14:933–934
- Cady WG (1964) Piezoelectricity, vol 2. Dover, New York, pp 699–711
- Camia FM (1967) Traité de thermocinétique impulsionnelle. Dunod, Paris
- Carslaw HJ, Jaeger JC (1959) Conduction of heat in solids, 2nd edn. Oxford University Press, New York
- Chen J, Zhang W, Feng Z, Cai W (2014) Determination of thermal contact conductance between thin metal sheets of battery tabs. *Int J Heat Mass Transf* 69:473–480
- Chynoweth AG (1956) Dynamic method for measuring the pyroelectric effect with special reference to barium titanate. *J Appl Phys* 27:78–84
- Clay W, Evans BJ, Latham RV (1974) A nondestructive pyroelectric display of an antiparallel polarization distribution in single-crystal barium titanate. *J Phys D Appl Phys* 7:1291–1295
- Cole KS, Cole RH (1941) Dispersion and absorption in dielectrics I. Alternating current characteristics. *J Chem Phys* 9:341–351
- Collins RE (1977) Measurement of charge distribution in electrets. *Rev Sci Instrum* 48:83–91
- Cook WR, Berlincourt DA, Scholz FJ (1963) Thermal expansion and piezoelectricity in lead titanate zirconate and barium titanate. *J Appl Phys* 34:1392–1398
- Coufal H, Grygier RK (1989) Charge profiles of thin electret films mapped with subnanosecond thermal pulses. *J Opt Soc Am B* 6:2013–2017
- Coufal HJ, Grygier RK, Horne DE, Fromm JE (1987) Pyroelectric calorimeter for photothermal studies of thin films and adsorbates. *J Vac Sci Technol A* 5:2875–2889
- D’Azzo JJ, Houpis CH (1999) Nyquist, Bode and Nickols plots. In: Levine WS (ed) Control system fundamentals. CRC Press, Boca Raton, chapter 10.3
- Dagher G, Holé S, Lewiner J (2006) A preliminary study of space charge distribution measurements at nanometer spatial resolution. *IEEE Trans Dielectr Electr Insul* 13:1036–1041
- Davidson DW, Cole RH (1951) Dielectric relaxation in glycerol, propylene glycol, and n-propanol. *J Chem Phys* 19:1484–1490
- Devonshire AF (1954) Theory of ferroelectrics. *Adv Phys* 3:85–130
- Eydam E, Suchaneck G, Esslinger S, Schönecker A, Neumeister P, Gerlach G (2014) Polarization characterization of PZT disks and of embedded PZT plates by thermal wave methods. *AIP Conf Proc* 1627:31–36
- Eydam A, Suchaneck G, Schwankl M, Gerlach G, Singer RF, Körner C (2015) Evaluation of polarisation state of light metal embedded piezoelectrics. *Adv Appl Ceram* 114:226–230
- Eydam A, Suchaneck G, Gerlach G (2016a) Characterisation of the polarisation state of embedded piezoelectric transducers by thermal waves and thermal pulses. *J Sensor Sensor Syst* 5:165–170
- Eydam A, Suchaneck G, Gerlach G (2016b) Non-destructive evaluation of integrated piezoelectric transducers by thermal waves and thermal pulses. *Procedia Technol* 26:59–65

- Eydam A, Suchaneck G, Gerlach G (2016c) Thermal-pulse method for life monitoring of integrated piezoelectric transducers. *Procedia Eng* 168:848–851
- Flössel M, Gebhardt S, Schönecker A, Michaelis A (2010) Development of a novel sensor-actuator-module with ceramic multilayer technology. *J Ceram Sci Technol* 1:55–58
- Gaudenzi P (2009) Smart structures: physical behaviour, mathematical modelling and applications. Wiley, Chichester
- Gerlach G, Shvedov D, Norkus V (2004) Packaging influence on acceleration sensitivity of pyroelectric infrared detectors. In: Proceedings of sixth IEEE CPMT conference on high density microsystem design and packaging and component failure analysis, HDP'04. <https://doi.org/10.1109/HPD.2004.1346715>
- Gerlach G, Suchaneck G, Movchikova A, Malyshkina OA (2007) Nondestructive testing of ferroelectrics by thermal wave methods. *Proc SPIE* 6530:65300B. <https://doi.org/10.1117/12.715534>
- Glass AM (1968) Dielectric, thermal, and pyroelectric properties of ferroelectric LiTaO<sub>3</sub>. *Phys Rev* 172:564–571
- Göber H, Erk S, Grigull U (1955) Die Grundgesetze der Wärmeübertragung. Springer, Berlin, p 82
- Groetsch CW (2007) Integral equations of the first kind, inverse problems and regularization: a crash course. *J Phys Conf Ser* 73:012001
- Hadni A, Thomas R (1972) Laser study of reversible nucleation sites in triglycine sulphate and applications to pyroelectric detectors. *Ferroelectrics* 4:39–49
- Hadni A, Henninger Y, Thomas R, Vergnat P, Bruno Wyncke B (1965) Sur les propriétés pyroélectriques de quelques matériaux et leur application à la détection de l'infrarouge. *J Phys France* 26:345–360
- Hadni A, Thomas R, Perrin J (1969) Response of a triglycine sulphate pyroelectric detector to high frequencies (300 kHz). *J Appl Phys* 40:2740–2745
- Havriliak S, Negami S (1967) A complex plane representation of dielectric and mechanical relaxation processes in some polymers. *Polymer* 8:161–210
- Holé S (2008) Resolution of direct space charge distribution measurement methods. *IEEE Trans Dielectr Electr Insul* 15:861–871
- Hufenbach W, Gude M, Modler N, Heber T, Winkler A, Weber T (2013) Process chain modelling and analysis for the high-volume production of thermoplastic composites with embedded piezoceramic modules. *Smart Mater Res* 2013:201631
- Imburgia I, Romano P, Caruso M, Viola F, Miceli R, Riva Sanseverino E, Madonia A, Schettino G (2016) Contributed review: review of thermal methods for space charge measurement. *Rev Sci Instrum* 87:111501
- Kepler RG, Anderson RA (1978) Piezoelectricity and pyroelectricity in polyvinylidene fluoride. *J Appl Phys* 49:4490–4494
- Kremer F, Schoenhals A (eds) (2003) Broadband dielectric spectroscopy. Springer, Berlin, pp 62–72
- Landau LD, Lifshitz EM (1969) Mechanics (Volume 1 of a course of theoretical physics). Pergamon Press, Oxford, p 59
- Lang SB (1990) New theoretical analysis for the laser intensity modulation method (LIMM). *Ferroelectrics* 106:269–274
- Lang SB (1991) Laser intensity modulation method (LIMM): experimental techniques, theory and solution of the integral equation. *Ferroelectrics* 118:343–361
- Lang SB (1998) An analysis of the integral equation of the surface laser intensity modulation method using the constrained regularization method. *IEEE Trans Dielectr Electr Insul* 5:70–76
- Lang SB (2004) Laser intensity modulation method (LIMM): review of the fundamentals and a new method for data analysis. *IEEE Trans Dielectr Electr Insul* 11:3–12
- Lang SB (2006) Fredholm integral equation of the laser intensity modulation method (LIMM): solution with the polynomial regularization and L-curve methods. *J Mater Sci* 41:147–153

- Lang SB, Das-Gupta DK (1986) Laser-intensity-modulation method: a technique for determination of spatial distributions of polarization and space charge in polymer electrets. *J Appl Phys* 59:2151–2160
- Lang SB, Fleming R (2009) A comparison of three techniques for solving the Fredholm integral equation of the laser intensity modulation method (LIMM). *IEEE Trans Dielectr Electr Insul* 16:809–814
- Lang SB, Rosenman G, Rushin S, Kugel V, Nir D (1992) Electron emission and spontaneous polarization distribution of proton-exchanged LiNbO<sub>3</sub>. *Ferroelectrics* 133:253–258
- Leal Ferreira GF (1989) On the deconvolution of heat-pulse like signals. *J Appl Phys* 66:4924–4927
- Lines ME, Glass AM (1977) Principles and application of ferroelectrics and related materials. Clarendon Press, Oxford, chapter 5.2
- Liu ST, Zook JD (1974) Evaluation of Curie constants of ferroelectric crystals from pyroelectric response. *Ferroelectrics* 7:171–173
- Liu S, Grinberg I, Rappe AM (2013) Exploration of the intrinsic inertial response of ferroelectric domain walls via molecular dynamics simulations. *Appl Phys Lett* 103:232907
- Mah JW, Shammugan S, Mutharasu D (2017) Impact of temperature, pressure, and current on thermal resistance of thermal interface material in optoelectronics device. *J Optoelectron Biomed Mater* 9:79–84
- Malyshkina OV, Movchikova AA, Suchaneck G (2007) New method for the determination of the pyroelectric current spatial distribution in ferroelectric materials [in Russian]. *Phys Solid State* 49:2144–2147 [Fiz Tverd Tela 49:2045–2048]
- Malyshkina OV, Movchikova AA, Grechishkin RM, Kalugina ON (2010) Use of the thermal square-wave method to analyze polarization state in ferroelectric materials. *Ferroelectrics* 400:63–75
- Mandelis A, Zver MM (1985) Theory of photopyroelectric spectroscopy of solids. *J Appl Phys* 57:4421–4430
- Mandelis A, Nicolaides L, Yan C (2001) Structure and the reflectionless/refractionless nature of parabolic diffusion-wave fields. *Phys Rev Lett* 87:020801
- Mellinger A (2004) Unbiased iterative reconstruction of polarization and space charge profiles from thermal-wave experiments. *Meas Sci Technol* 15:1347–1353
- Mellinger A, Singh R, Gerhard-Multhaupt R (2005a) Fast thermal-pulse measurements of space-charge distributions in electret polymers. *Rev Sci Instrum* 76:013903
- Mellinger A, Singh R, Wegener M, Wirges W, Gerhard-Multhaupt R, Lang SB (2005b) Three-dimensional mapping of polarization profiles with thermal pulses. *Appl Phys Lett* 86:082903
- Mellinger A, Flores-Suárez R, Wegener M, Wirges W, Gerhard-Multhaupt R, Singh R (2006) Thermal-pulse tomography of polarization distributions in a cylindrical geometry. *IEEE Trans Dielectr Electr Insul* 13:1030–1035
- Mopsik FI, DeReggi AS (1982) Numerical evaluation on the dielectric polarization distribution from thermal pulse data. *J Appl Phys* 53:4333–4339
- Movchikova A, Malyshkina OV, Pedko BB, Suchaneck G, Gerlach G (2008a) Polarization profiling of ferroelectrics by thermal square wave methods. *Ferroelectrics* 367:38–44
- Movchikova A, Malyshkina O, Suchaneck G, Gerlach G, Steinhausen R, Langhammer HT, Pentschke C, Beige H (2008b) Study of the pyroelectric behavior of BaTi<sub>1-x</sub>Sn<sub>x</sub>O<sub>3</sub> piezoceramics. *J Electroceram* 20:43–46
- Movchikova A, Malyshkina OV, Pedko BB, Suchaneck G, Gerlach G (2009) The influence of doping on the pyroelectric response of SBN single crystals. *Ferroelectrics* 378:186–194
- Movchikova A, Suchaneck G, Malyshkina OV, Pedko BB, Gerlach G (2010) Thermal wave study of piezoelectric coefficient distribution in PMN-PT single crystals. *Adv Appl Ceram* 109:131–134
- Neugschwandner GS, Schwödiauer R, Bauer-Gogonea S, Bauer S (2001) Piezo- and pyroelectricity of a polymer-foam space-charge electret. *J Appl Phys* 89:4503–4511

- Notingher P, Agnel S, Fruchier O, Toureille A, Rousset B, Sanchez JL (2004) On the use of the thermal step method as a tool for characterizing thin layers and structures for micro and nanoelectronics. *J Optoelectron Adv Mater* 6:1089–1906
- Notingher P, Holé S, Baudon S, Fuchier O, Boyer L, Agnel S (2012) Toward non-destructive high resolution thermal methods for electric charge measurements in solid dielectrics and components. In: ESA2012- electrostatics joint conference, Cambridge, June 2012. Online publication. [http://www.electrostatics.org/images/ESA2012\\_M3.pdf](http://www.electrostatics.org/images/ESA2012_M3.pdf). Accessed 24 Nov 2017
- Nye JF (1995) Physical properties of crystals: their representation by tensors and matrices. Clarendon Press, Oxford, chapter 10.4.4
- Parker WJ, Jenkins RJ, Butler CP, Abbott GL (1961) Flash method of determining thermal diffusivity, heat capacity, and thermal conductivity. *J Appl Phys* 32:1679–1684
- Perrin B, Bonello B, Jeannet J-C, Romatet E (1996) Interferometric detection of hypersound waves in modulated structures. *Prog Nat Sci* 6:444–448
- Perry CH, Khan BN, Rupprecht G (1964) Infrared studies of perovskite titanates. *Phys Rev* 135: A408–A412
- Peterson RL, Day GW, Gruzensky PM, Phelan RJ (1974) Analysis of response of pyroelectric optical detector. *J Appl Phys* 45:3296–3303
- Petre A, Marty-Dessus D, Berquez L, Franceschi JL (2006) Space charge cartography by FLIMM on SEM-irradiated PTFE thin films. *J Electrost* 64:492–497
- Petzelt J, Nuzhnyy D, Bovtun V, Kempa M, Savinov M, Kamba S, Hlinka J (2015) Lattice dynamics and dielectric spectroscopy of BZT and NBT lead-free perovskite relaxors – comparison with lead-based relaxors. *Phase Transit* 88:320–332
- PI Ceramic GmbH (2017) Piezoelectric ceramic products: fundamentals, characteristics and applications. <https://www.piceramic.com/en/products/piezoceramic-materials/#c15193>. Accessed 24 Nov 2017
- Ploss B, Bianzano O (1994) Polarization profiling of the surface region of PVDF and P(VDF-TrFE). In: IEEE 8th international symposium on electrets. <https://doi.org/10.1109/ISE.1994.514769>
- Ploss B, Emmerich R, Bauer S (1992) Thermal wave probing of pyroelectric distributions in the surface region of ferroelectric materials: a new method of analysis. *J Appl Phys* 72:5363–5370
- Prudnikov AP, Brychkov YA, Marichev OI (1992) Integrals and series, vol. 1: Elementary functions. Taylor & Francis, London, p 730
- Putley EH (1970) The pyroelectric detector. In: Willardson RK, Beer AC (eds) Semiconductors and semimetals, vol 5. Academic, New York, pp 259–285
- Qiu X, Hollander L, Flores Suarez R, Wirges W, Gerhard R (2010) Polarization from dielectric-barrier discharges in ferroelectrets: mapping of the electric-field profiles by means of thermal-pulse tomography. *Appl Phys Lett* 97:072905
- Raffmetal (2017) EN 46000 data sheet. [www.raffmetal.com/scarica\\_file.asp?c=/dati/SearchAlloy/ENG/&f=EN46000.pdf](http://www.raffmetal.com/scarica_file.asp?c=/dati/SearchAlloy/ENG/&f=EN46000.pdf). Accessed 24 Nov 2017
- Reboul JM (2011) Thermal waves interferences for space charge measurements in dielectrics. In: 14th international symposium on electrets (ISE). <https://doi.org/10.1109/ISE.2011.6084983>
- Reboul JM, Mady F (2004) Space charge measurements by the alternating thermal wave method: thermal analysis and simulations for data processing improvement. In: Proceedings of the 2004 I.E. international conference on solid dielectrics ICSD. <https://doi.org/10.1109/ICSD.2004.1350339>
- Reboul JM, Cherifi A, Carin R (2001) A new method for space charge measurements in dielectric films for power capacitors. *IEEE Trans Dielectr Electr Insul* 8:753–759
- Rose A (1973) Vision – human and electronic. Plenum Press, New York
- Rübner M, Körner C, Singer RF (2008) Integration of piezoceramic modules into die castings – procedure and functionalities. *Adv Sci Technol* 56:170–175
- Sakai S, Date M, Furukawa T (2002) Development of polarization distribution in fatigued films of ferroelectric vinylidene fluoride/trifluoroethylene copolymer. *Jpn J Appl Phys* 41:3822–3828
- Salazar A (2006) Energy propagation of thermal waves. *Eur J Phys* 27:1349–1355
- Samoilov VB, Yoon YS (1998) Frequency response of multilayer pyroelectric sensors. *IEEE Trans Ultrason Ferroelectr Freq Control* 45:1246–1254

- Sandner T (2003) Verfahren zur tiefenaufgelösten Polarisationsbestimmung von pyroelektrischen PZT-Dünnsschichten für IR-Sensoren. w.e.b.-Universitätsverlag, Dresden
- Sandner T, Suchaneck G, Koehler R, Suchaneck A, Gerlach G (2002) High frequency LIMM – a powerful tool for ferroelectric thin film characterization. *Integr Ferroelectr* 46:243–257
- Schein LB, Cressman PJ, Cross LE (1978) Electrostatic measurements of tertiary pyroelectricity in partially clamped LiNbO<sub>3</sub>. *Ferroelectrics* 22:945–948
- Sea Ceramics Technology (2017a) Thermal properties of ceramics and metal/metal composites. <http://seaceramics.com/Download/Reference%20Data/Cer-Metal%20thermal%20properties.pdf>. Accessed 24 Nov 2017
- Sea Ceramics Technology (2017b) Table of LTCC materials properties. <http://seaceramics.com/Download/Reference%20Data/LTCC%20PROPERTY2.pdf>. Accessed 24 Nov 2017
- Sessler GM (1997) Charge distribution and transport in polymers. *IEEE Trans Dielectr Electr Insul* 4:614–628
- Stewart M, Cain M (2008) Spatial characterization of piezoelectric materials using the scanning laser intensity modulation method (LIMM). *J Am Ceram Soc* 91:2176–2181
- Suchaneck G, Lin W-M, Koehler R, Sandner T, Gerlach G, Krawietz R, Pompe W, Deineka A, Jastrabik L (2002) Characterization of RF-sputtered self-polarized PZT thin films for IR sensor arrays. *Vacuum* 66:473–478
- Suchaneck G, Hu W, Gerlach G, Flössel M, Gebhardt S, Schönecker A (2011) Nondestructive evaluation of polarization in LTCC/PZT piezoelectric modules by thermal wave methods. *Ferroelectrics* 420:25–29
- Suchaneck G, Eydam A, Hu W, Kranz B, Drossel W-G, Gerlach G (2012) Evaluation of polarization of embedded piezoelectrics by the thermal wave method. *IEEE Trans Ultrason Ferroel Freq Control* 59:1950–1954
- Suchaneck G, Eydam A, Rübner R, Schwankl M, Gerlach G (2013a) A simple thermal wave method for the evaluation of the polarization state of embedded piezoceramics. *Ceram Int* 39:S1:S587–S590
- Suchaneck G, Eydam A, Gerlach G (2013b) A laser intensity modulation method for the evaluation of the polarization state of embedded piezoceramics. *Ferroelectrics* 453:127–132
- Toureille A, Reboul JP (1988) The thermal-step-technique applied to the study of charge decay in polyethylene thermoelectrets. In: IEEE 6th international symposium on electrets (ISE 6). <https://doi.org/10.1109/ISE.1988.38518>
- Tuncer E, Lang SB (2006) Kramers-Kronig relations in laser intensity modulation method. *Phys Rev B* 74:113109
- van der Ziel A (1973) Pyroelectric response and  $D^*$  of thin pyroelectric films on a substrate. *J Appl Phys* 44:546–549
- Voigt W (1910) Lehrbuch der Kristallphysik. Teubner, Leipzig
- von Laue M (1925) Piezoelektrisch erzwungene Schwingungen von Quarzstäben. *Z Phys* 34:347–361
- von Seggern H (1978) Thermal-pulse technique for determining charge distributions: effect of measurement accuracy. *Appl Phys Lett* 33:134–137
- von Seggern H, West JE, Kubli RA (1984) Determination of charge centroids in two-side metallized electrets. *Rev Sci Instrum* 55:964967
- Wilkie WK, Bryant GR, High JW, Fox RL, Hellbaum RF, Jalink A, Little BD, Mirick PH (2000) Low-cost piezocomposite actuator for structural control applications. In: Proceedings of SPIE 3991, smart structures and materials 2000: industrial and commercial applications of smart structures technologies. <https://doi.org/10.1117/12.388175>
- Wu Q, Zhang X-C (1995) Free-space electro-optic sampling of terahertz beams. *Appl Phys Lett* 67:3523–3525
- Xu-Sheng W (1993) Tertiary pyroelectric effect on thick ferroelectric crystal plates with partially uniform heating. *Ferroelectr Lett Sect* 15:159–165

- Yilmaz S, Bauer S, Wirges W, Gerhard-Multhaupt R (1993) Scanning electro-optical and pyroelectrical microscopy for the investigation of polarization patterns in poled polymers. *Appl Phys Lett* 63:1724–1726
- Zajosz J (1979) Pyroelectric response to step radiation signals in thin ferroelectric films on a substrate. *Thin Solid Films* 62:229–236
- Zheng F, Zhang Y, An Z, Liu C, Dong J, Lin C (2013) Thermal pulse method with an applied field. In: IEEE international conference on solid dielectrics (ICSD). <https://doi.org/10.1109/ICSD.2013.6619841>
- Zook JD, Liu ST (1978) Pyroelectric effect in thin film. *J Appl Phys* 49:4604–4606



# Sonothermic Techniques in Nondestructive Evaluation 40

Xiaoyan Han

## Contents

Introduction .....	1480
Experimental Setup .....	1482
Applications of Sonic IR Imaging NDE for Defect Detection .....	1483
Applications on Materials and Structures with Welding .....	1484
Sonic IR Detection of Fatigue Cracks on Super Alloy Turbine Blades .....	1485
Fast, Wide-Area Inspection of Composite Structures .....	1486
Sonic IR Crack Detection in Civil Structures .....	1487
Defect Detection in Concrete Structures .....	1487
Quantitative Evaluation of Defect Depth by Sonic IR Imaging .....	1488
Nonlinearity in Sonic IR Imaging .....	1492
Summary .....	1494
References .....	1495

## Abstract

Sonic Infrared (IR) imaging is by far the world most sensitive and fastest crack detection technique. It is a novel hybrid sensing technology that utilizes the concept of combining infrared imaging with sonic/ultrasonic excitation. It can detect closed hairline cracks in safety critical materials and structures such as airplanes. Sonic IR imaging is among the youngest members of the NDE techniques. It can be employed to detect defects in materials and structures, such as cracks, delaminations, and disbonds for both surface and subsurface features. The ultrasonic excitation is typically a fraction-of-a-second-long pulse, which causes friction heating in the defects. The temperature changes in the target due to the heating are simultaneously being imaged by an infrared camera. Thus, defects in a

---

X. Han (✉)

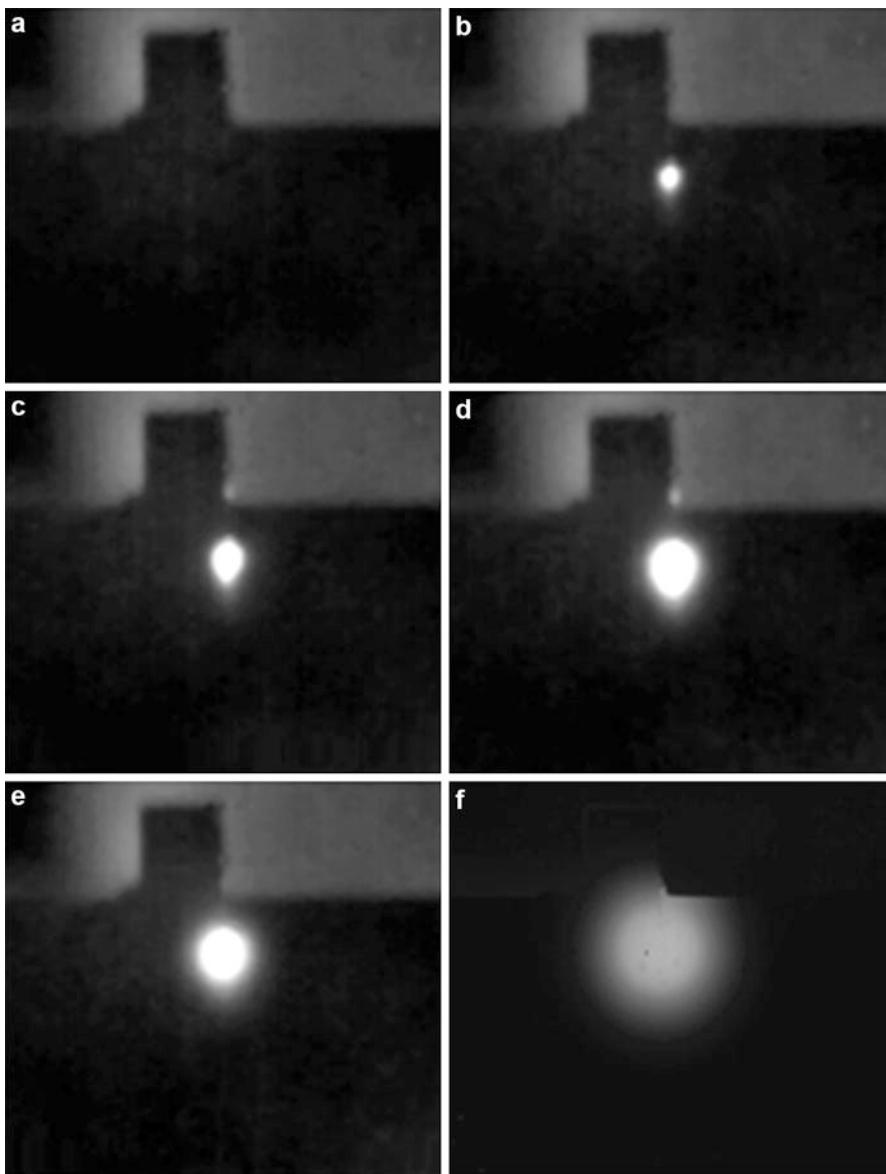
Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI, USA  
e-mail: [Xiaoyan.han@wayne.edu](mailto:Xiaoyan.han@wayne.edu)

wide variety of materials can be identified. This method is a fast, wide-area, and true dark field imaging NDE technique.

---

## Introduction

Sonic Infrared (IR) imaging, a novel hybrid sensing technology that utilizes the concept of combining IR imaging with sonic/ultrasonic excitation, is by far the world most sensitive and fastest crack detection technique. Based on the nature of the technology, it can detect a variety of defects at different locations, such as cracks, delaminations, and disbonds located either surface or subsurface in materials and structures, including closed hairline cracks in critical materials and structures such as airplanes (Henneke et al. 1979; Reifsneider et al. 1980; Mignogna et al. 1981; Wu et al. 1996; Favro et al. 2000, 2001a, b; Han et al. 2001, 2002a; Thomas et al. 2001). Sonic IR imaging is among the youngest members of the NDE techniques. Typical NDE techniques include X-Ray, (Bray and Stanley 1997a) ultrasouinic, (Metals Handbook 1989; Bray and Stanley 1997b) eddy current, (Bray and Stanley 1997c; Phillips 1998) magnetic flux leakage technique, (Bray and Stanley 1997d; Betz 1997) fluorescent penetration inspection (Bray and Stanley 1997e; Lovejoy 1991). Sonic IR imaging has by far the broadest applications comparing with the traditional NDE/NDI/NDT techniques since it can be applied to ferrous and nonferrous materials, electrically conductive and electrically nonconductive materials, metal/metal alloy or non-metal materials, single materials or composite materials, regular geometry or nonregular geometry objects. The ultrasonic excitation pulse is typically a fraction of a second in length. The sound vibrations cause relative motions between faying faces when the sound waves propagate within the target object, which produce localized frictional heating or other irreversible internal surface interactions in the vicinity of defects. The temperature changes in the target due to this heating are being imaged simultaneously by an infrared camera. If the local heating at the defects cause large enough temperature change on the target surface to be detected by an IR camera, the defects can then be identified. If the defects intersect the surface or very close to the surface where it is imaged, then the heat sources first appear as patterns with clear outlines of the defects in the IR image. These outlines subsequently get broader and more blurred into surrounding regions due to heat diffusion. If the defects are located deeper under the surface, then the heat sources appear as smeared heat patterns. In general, localized heating accumulates at defects during the period of the excitation pulse. However, when the sound pulse is turned off, the resulting heat spreads through the thermal diffusion process, and without energy supply, the material reaches thermal equilibrium eventually. The whole imaging period takes only a second or so for metal objects, and maybe a few seconds for materials with lower thermal conductivity, or for the cases that defects are buried deep in the target. Figure 1 shows a sequence of IR images of roughly one millimeter long crack in a titanium sample. As one can see, the crack blooms in a burst in Sonic IR imaging. These images were taken at different ultrasound excitation stages: (a) before the excitation source was on; (b-d) during the 0.2-sec



**Fig. 1** A sequence of IR images of roughly 1-mm long crack in a titanium sample blooms like a bomb burst in Sonic IR imaging. The IR images were taken at different ultrasound excitation stages: (a) before the pulse; (b-d) during the pulse, (e) shortly after the pulse, and (f) about 1 s after the pulse

excitation, (e) shortly after the ultrasound excitation was off, and (f) about 1 s after the excitation was off. The heating mechanism and process are very different from that of thermoelastic heating and cooling. The latter is a reversible process and has no net heating effect because it would be averaged out during the typical IR sensor's

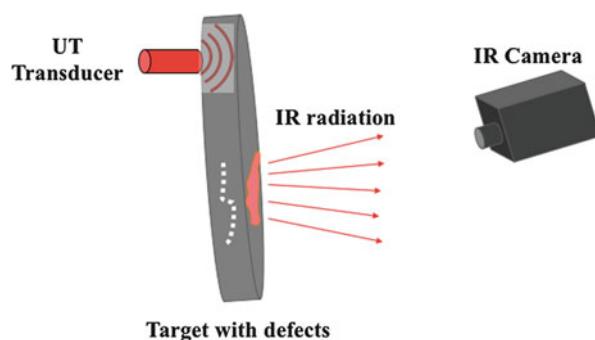
integration time. For the purpose of doing stress pattern analysis using thermal emission, fast thermal detectors have to be used to catch the reversible heating and cooling patterns. In the sonic IR imaging approach, only the irreversible heating is imaged by the IR camera. Since it is at the defects where local heating occurs, the Sonic IR method is not only a fast, wide-area technique but also a true dark field imaging NDE method.

## Experimental Setup

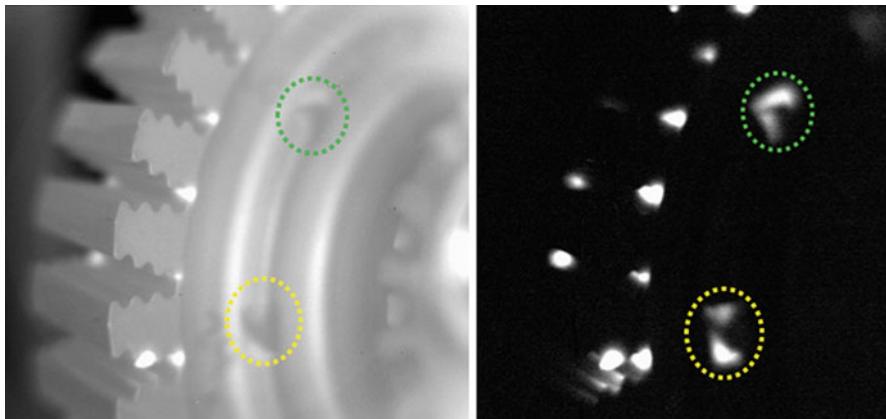
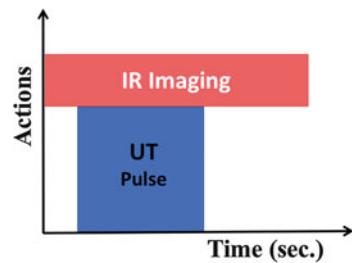
The schematic drawing in Fig. 2 illustrates the key components in a typical Sonic IR imaging evaluation. An ultrasonic excitation source is applied to the sample to cause friction heating in the defects; an infrared camera is used to image the resulting heating from the defect or defects in the target sample. Since ultrasound travels in solid materials at a few thousand meters per second, therefore, the location of the transducer is usually not an issue, except to choose a convenient spot. In many practical objects' inspections/evaluations, the IR camera would be placed on the same side as the ultrasound transducer due to the inaccessibility of the other side of the target structures. Of course, the proper way is to have the camera image the regions of interest in the target.

The current Sonic IR imaging system is configured such that an IR camera is turned on to acquire images a short time before the ultrasound source is turned on to inject energy into the sample, and the IR camera is kept on to acquire images for a certain time even after the end of the ultrasound pulse. This scheme is shown in Fig. 3 to illustrate the timing relationship between the ultrasound pulse and the IR image acquiring process. The images taken prior to the ultrasound pulse can be regarded as background and can be subtracted from any image among the image sequence. The resulting sequence of images provides the contrast data, typically with dark background. These dark background contrast images usually have reduced background artifacts, including infrared reflection from background objects; therefore, this process can greatly increase the sensitivity of this technology for defect detection. In Fig. 4, a raw IR image and a contrast IR image of a turbine disk are presented to show the dramatic effect, the former does give confident call for *some*

**Fig. 2** Schematic drawing of a typical Sonic IR imaging arrangement



**Fig. 3** Schematic drawing of timing relationship between the ultrasound excitation pulse and IR imaging



**Fig. 4** Comparison of a raw IR image and a contrast IR image of a turbine disk

cracks in the turbine disk, but the latter one provides definite answer for these cracks. Furthermore, the latter one also shows the two groups of defects, indicated by the blue (upper) and yellow (lower) dashed ellipses, which do not give obvious indications of the existence of the defects at those locations with the corresponding colors of dashed line ellipses.

### Applications of Sonic IR Imaging NDE for Defect Detection

Sonic IR Imaging imaging is the youngest member among the NDE family which include X-Ray, ultrasonic, eddy current, magnetic flux leakage technique, and fluorescent penetration inspection. Sonic IR imaging has proved its broad application scope as well. It basically has no limitations on the types of materials, and works on ferrous and non-ferrous materials, electrically conductive and electrically non-conductive materials, metal/metal alloy or nonmetal materials, single materials or composite materials, regular geometry or nonregular geometry objects. A few examples are presented here to demonstrate some possible applications of Sonic IR imaging NDE.

## Applications on Materials and Structures with Welding

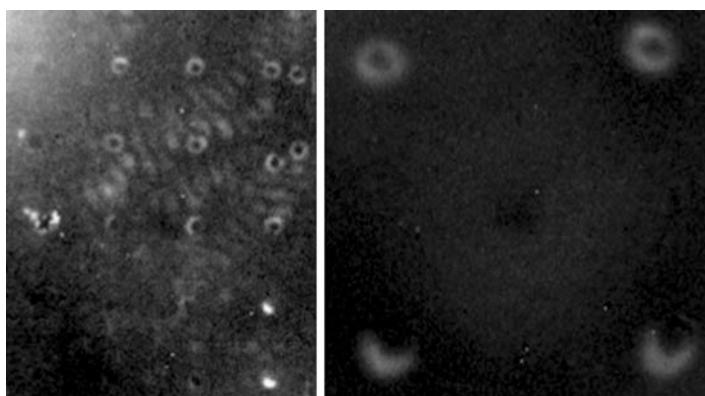
Welding is a common fabrication process to join materials together, same type or different types. There are many different welding techniques, but none of them guarantees 100% quality assurance. Here are a couple of examples to using sonic IR imaging as a potential method to inspect welds.

### 1. Sonic IR Imaging Evaluation of Spot Welds for Product Quality Control

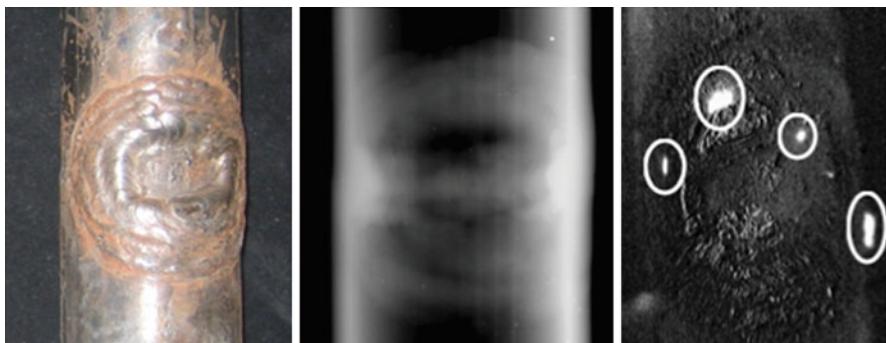
Spot welding is a common process for joining two metal structures, often sheets, together by concentrating welding current into a small “spot” and thus melting the metal at the spot to simultaneously clamp the metal together. Spot welding is a fast process which could only take as little as tens of a millisecond. It is therefore commonly used in manufacture, such as in the automobile industry. However, bad welds occur, which of course have an effect on the structure and product quality. Sonic IR imaging can be a potential tool for this problem. In Fig. 5, two IR images are shown: The image on the left presents wide area inspection of welded aluminum panels which have over 20 spot welds; the image on the right, a zoomed-in image, shows better details and high resolution of four of the spot welds. In the right image, the good spot welds, the upper two in the right image, are very well distinguish them from the bad spot welds, the lower two in the right image: Good spot welds are shown in perfect circles, and the bad ones are shown in partial circles. Of course, the spot welds are located under the top sheet, thus, this is an application of subsurface feature evaluation.

### 2. Detect Disbands and Cracks in Welded Steel Power Facilities

Some preliminary experiments have been done on steel boiler tubes provided by an energy company who was seeking for an effective, easy, fast, and unambiguous NDE technology to detect possible disbonds and cracks in boiler tube windows with welds. A picture of a sample boiler tube containing a window weld is shown in the left image in Fig. 6. The sample tube has outer



**Fig. 5** Sonic IR images of spot welds



**Fig. 6** Optical image (left), X-ray image (center), and Sonic IR image (right) of a nuclear power steel boiler tube with a welded window area

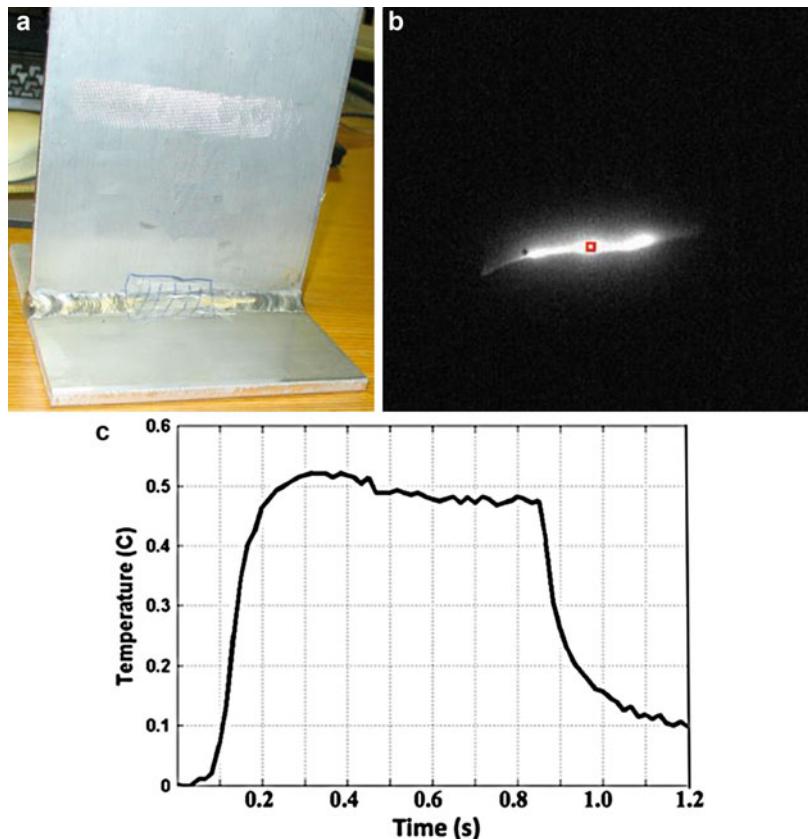
diameter of 50 mm and inner diameter of 38 mm with wall thickness of 6 mm. The X-ray image at the center does not show clear identification of cracks in the tube ([X-ray](#) image). Sonic IR imaging, however, indicates that there are defects in the tube as marked by white circles on the right image. Since there are no visual signs of cracks or disbonds on the surface, the defects in the IR image are subsurface. The sample tube was later destructively examined; the results from the examination confirmed the sonic IR findings in the tube window area.

### 3. Defect Detection on T-Joint Structures: Subsurface

T-joint is one type of geometrical shapes commonly used in many structures such as ships. Figure 7a shows a T-joint sample with two 153 mm × 153 mm × 6.4 mm aluminum plate welded together. A crack was detected along the center part of the welded area as marked, and it is shown in the IR image in Fig. 7b. The IR signal of a spot, marked as a small red square in the center image, is plotted to show the IR signal during the 0.8-sec ultrasound pulse. Both the IR image and the temperature-time plot can be used as signatures in flaw identification.

### Sonic IR Detection of Fatigue Cracks on Super Alloy Turbine Blades

Turbine blades make up the essential and critical elements in both gas turbines and steam turbine engines. Blade fatigue is a major source of failure in turbines. Typically, turbine blades are subjected to very strenuous environments inside a turbine engine, for example, high temperature, high stresses due to a high rotation speed, and/or high vibration. Although turbine blades are typically made of super alloy to give greater strength and/or resistance to corrosion, these factors can still lead to blade failures, and as a result to potentially destroy the engine. Therefore, effective inspection of turbine blades is critical to prevent fatal accidents. Figure 8 shows one IR image of an aircraft turbine engine blade. Multiple cracks around the edges of the blade are identified by the sonic IR technique.



**Fig. 7** (a) Photo of the T-joint of two Al plates welded together; (b) Sonic IR image of a crack along the center weld area; (c) Temperature-time plot of the spot marked by the red square in the IR image

### Fast, Wide-Area Inspection of Composite Structures

It has been demonstrated that sonic IR imaging is also an excellent technology for detecting disband/delamination in composite materials, such as in advanced carbon-fiber composites, especially on “kissing disbonds” while other techniques have their limitations on such type of defects (Mian et al. 2004; Han 2007; Han et al. 2010a, b). Experiments have shown that sonic IR imaging has capability of accurate and fast inspection for wide areas. A composite panel of Airbus 330 vertical stabilizer is about 1530 mm long and 610 mm wide. Its skin thickness is 4.1 mm, and the thickness over the rim is 5.4 mm. The photo of the panel is shown on the left side in Fig. 9, and the IR image of the whole panel is shown on the right. The sonic IR image clearly shows the backside features as circled in red. It demonstrates that sonic IR not only can “see through” this thickness of the panel but also can inspect a big area of such composite structures within a very short time, only a few seconds in this case.

**Fig. 8** IR image of multiple cracks around the edges of a turbine blade

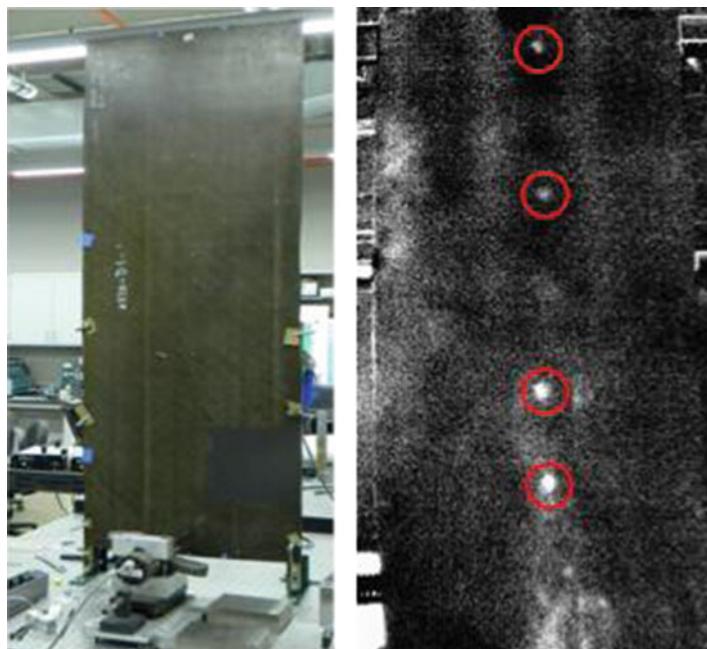


### Sonic IR Crack Detection in Civil Structures

Many civil structures, such as buildings and bridges, use steel C-channel beams. With aging and bearing continuous high stress, such steel structures can grow fatigue cracks as well. A drawing of the side view and top view of such a C-channel beam with its dimensions is shown in Fig. 10. The beam is 840 mm × 125 mm × 48 mm. It was subjected to thermal fatigue. Figure 11 shows a photo of a smaller cut piece on the left, and a sonic IR image of a group of cracks in the C-channel beam on the right. The IR image is a contrast one with the background information subtracted to produce a high signal to noise ratio image.

### Defect Detection in Concrete Structures

There has been a limited research work done on concrete structures with sonic IR imaging. It takes much longer time for heat to diffuse in concrete structures because of the thermal properties of such materials. Therefore, any defects beneath the surface could take seconds or even minutes to be detected by the thermal technique depending on how deep the defects are located under the surface. Despite the thermally slow nature of concrete, defect detection using sonic IR imaging is still doable. A photo of a cylindrical concrete pillar with a diameter of 160 mm and height of 320 mm is shown in Fig. 12. The surface has many marks and pits, as shown in the photo on the left side. The sonic IR image on the right shows subsurface fracture in the upper part of the pillar. The pits on the surface do not show up in the contrast IR image as expected.



**Fig. 9** Left: Photo of the A330 1530 mm × 610 mm composite panel; Right: Sonic IR image demonstrates its capability of accurate and fast inspection for large areas

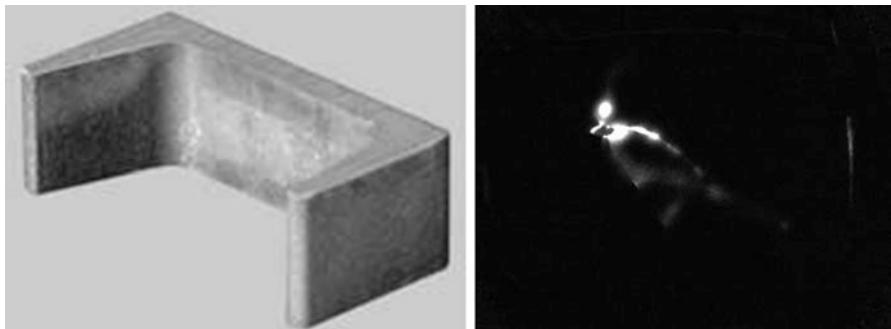


**Fig. 10** Dimensions of a steel C-channel labeled on the drawings of its side view and top view

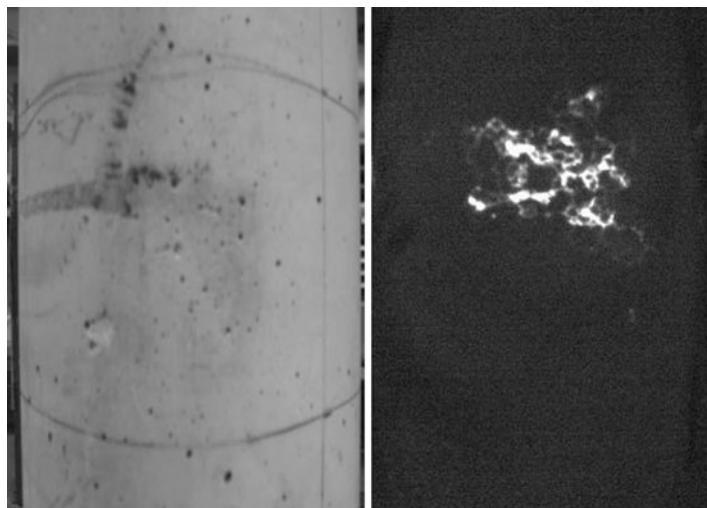
In addition to the above listed applications, there are other problems to which the use of Sonic IR imaging can be applicable, such as corrosion detection (Han 2007).

### Quantitative Evaluation of Defect Depth by Sonic IR Imaging

Sonic IR imaging technique provides infrared signal that can be used to identify the problems in the target. In the example below shown in Fig. 13, two Sonic IR images were taken at the times when the signals from the crack reached its peak levels in two experiments under different conditions, and their corresponding IR signal vs. time



**Fig. 11** Left: A photo of a smaller cut piece; Right: A Sonic IR image of a group of cracks in the C-channel beam

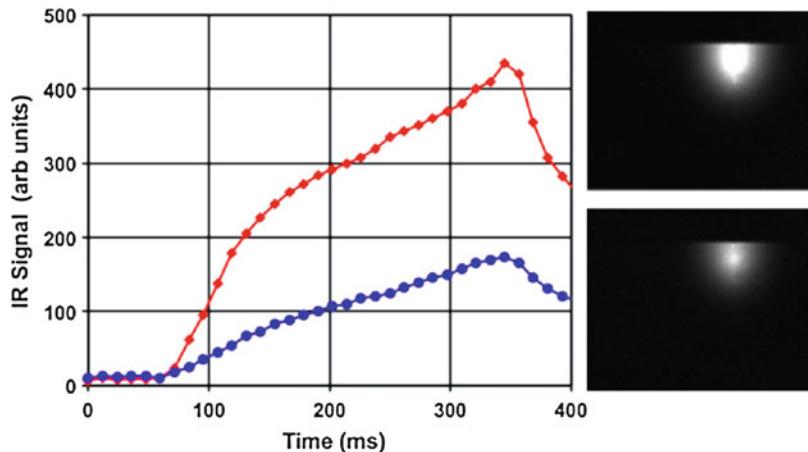


**Fig. 12** Left: A photo of a cylindrical concrete pillar with surface pits and marks; Right: A Sonic IR image of subsurface fracture in the upper part of the pillar

curves are plotted as well. Both the plots and the images demonstrate that these IR signatures can distinguish the two experimental results easily.

IR signal data analysis methods can be used to extract physical quantities for quantitative measurement in applications. In Sonic IR imaging, some of the input ultrasound energy is converted to heat at the defects. Thus, the defects themselves act as heat sources. Heat diffusion from the heat sources to their vicinity obeys the physics law of thermodynamics. The heat equation, which describes the heat distribution or temperature variation in a particular body over time, can be written as:

$$\rho c \frac{\partial T}{\partial t} - \nabla(\kappa \nabla T) = q$$



**Fig. 13** IR signatures of a crack from two different experimental settings: IR signals at the defect as functions of time (left), and the IR images taken at the time of peak IR signal levels (right)

Where  $T(x, y, z, t)$  is the Temperature in Kelvin;  $t$  is time;  $\rho$  is the mass density in  $\text{kg m}^{-3}$ ;  $c$  is the specific heat capacity in  $\text{J K}^{-1} \text{kg}^{-1}$ ;  $k$  is the thermal conductivity in  $\text{W m}^{-1} \text{K}^{-1}$ ;  $q(x, y, z, t)$  is the heat source in W;  $\nabla$  is the first order partial differential operator with respect to the three spatial directions  $x$ ,  $y$ , and  $z$ .

Different cases have different material properties and boundary conditions. For instance, metal and metal alloy materials typically are good thermal conductors and have higher thermal conductivity compares to ceramics and plastics. Many composite materials have inhomogeneous and anisotropic structures. For this type of materials, the mass density, thermal conductivity, and specific heat capacity vary from location to location within the structure. Heat equations can be derived for the individual cases based on their material properties and their corresponding boundary conditions. As an example, layered carbon fiber-reinforced composite structures are used in which multiple layers are laminated together. It is a good approximation to make assumptions of homogeneous and isotropic conditions in the in-plane dimensions and anisotropic conditions in the out-of-plane dimension. The mathematical expression of the surface temperature ( $T$ ) versus time ( $t$ ) for these materials is shown in the equation below (Han et al. 2012):

$$T(\vec{0}, t) = d \sum_{n=-\infty}^{+\infty} \left[ t^{\frac{1}{2}} \times e^{-\frac{g}{t}} - \sqrt{g\pi} + \sqrt{g\pi} \times \operatorname{erf}\left(\sqrt{\frac{g}{t}}\right) \right] - d \sum_{n=-\infty}^{+\infty} \left[ t^{\frac{1}{2}} \times e^{-\frac{f}{t}} - \sqrt{f\pi} + \sqrt{f\pi} \times \operatorname{erf}\left(\sqrt{\frac{f}{t}}\right) \right].$$

In which

$$d = \frac{16\rho c \pi \alpha_1}{(4\pi\alpha_1\alpha_2\alpha_3)^{\frac{3}{2}}}.$$

$$g = \frac{(b-2na)^2}{4\alpha_3},$$

$$f = \frac{\alpha_3 R^2 + (b-2na)^2 \alpha_1}{4\alpha_1 \alpha_3}$$

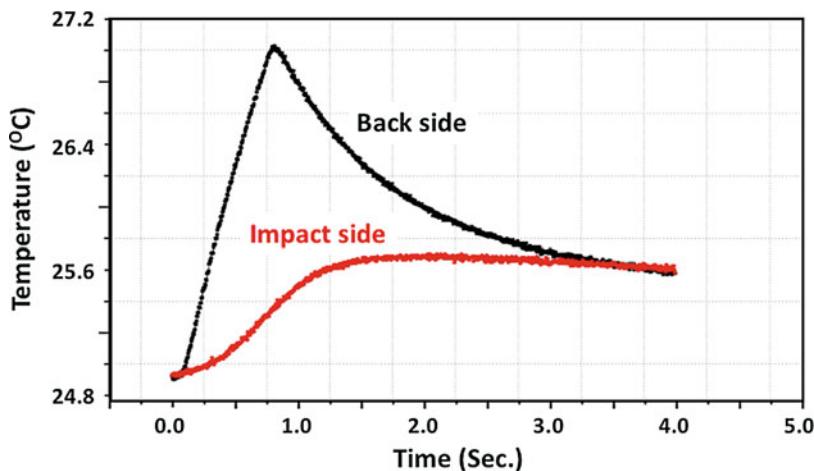
$$\alpha_i \equiv \frac{\kappa_i}{\rho_i c_i}, \quad i = 1, 2, 3$$

where  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  are the thermal diffusivities of the carbon fiber-reinforced composite plate in the  $x$ ,  $y$ , and  $z$  dimensions; parameter “ $b$ ” represents the thickness of the composite plate; parameter “ $a$ ” represents the depth of a delamination defect below the inspection surface; and  $R$  represents the delamination lateral size. With the measured temperature-time curve(s) and the thermal properties of the material, the defect depth can be calculated.

A carbon fiber-reinforced composite sample with dimensions 305 mm  $\times$  305 mm and eight-ply thick was subjected to an impact which caused delaminations between plies at different depths. Both sides, the impact side and the back side, are inspected. Two images of the two sides are shown in Fig. 14. IR signals of the corresponding spot of the two sides, indicated by the dark circle (left) and the red circle (right), were extracted and converted to temperatures. The temperature-time data are plotted and shown in Fig. 15. One can see that the signal levels from the corresponding spot on the two sides are different, with the backside signal much higher than the impact side. Furthermore, the two signals reach their peaks at different times, with the backside signal reaching its peak at earlier time of 0.8 s than the impact signal at time of 2.15 s. This situation is quite common for impact damage in which shock waves cause damages closer to the farther end from the impact location. These temperature-time curves can be used to derive the defect depth information.



**Fig. 14** Sonic IR images of both sides of the carbon fiber-reinforced composite panel which was subjected to an impact damage



**Fig. 15** Temperature-time plots on the corresponding spot of the two sides, indicated by the dark circle and the red circle in Fig. 14

### Nonlinearity in Sonic IR Imaging

Nonlinearity is an important phenomenon which occurs in the coupling between the ultrasound transducer and the target in Sonic IR imaging under certain conditions. This nonlinearity refers to the produced *multiple* frequencies of vibration in the sample from an input of *single* frequency ultrasound source. This phenomenon was first discovered when sound was heard during the ultrasound pulse from a *single* driving frequency source with 20 kHz, or 40 kHz which are not audible. The produced multiple frequencies could be a collection of the harmonics of the driving frequency, or could be a collection of integral multiples of a fraction of the driving frequency of the excitation source (Han et al. 2002b). Therefore, the vibration waveforms in the sample are no longer a single frequency regular sinusoidal vibration as that of the input pulse but rather have very irregular waveforms. The lowest frequency in the spectrum of the irregular waveforms could go as low as about 1 kHz and as high as over 240 kHz, while the driving frequency from the ultrasonic source is 20 kHz, or 30 kHz, or 40 kHz, or some other frequency. This complicated acoustic behavior was given the name of “acoustic chaos.” Both analytical modeling and finite element modeling proved the natural occurrence of acoustic chaos in Sonic IR imaging with its high sensitivity to the prior condition at any moment in the process (Han et al. 2004a, 2005, 2006). Further studies have demonstrated that acoustic chaos actually enhances defect detectability and thus improve probability of detection (POD) of Sonic IR imaging NDE (Han et al. 2004b; Song and Han 2013a, b). The two Sonic IR images in Fig. 13 were taken from two experiments with different conditions: the upper one under chaotic excitation, and the lower one under non-chaotic condition, and both images were taken at their peak

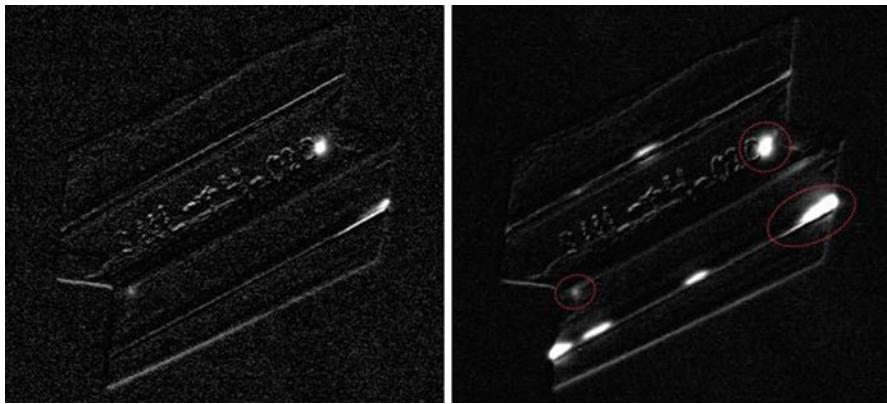
signal level at the crack. Although the crack can be detected under either of the conditions for this sample, the very different signal levels are a giveaway of the preferred situation.

In other cases, however, chaotic excitation may be the only choice for defect detection. A photo of a commercial turbine blade is shown in Fig. 16. The blade dimensions are 222 mm in length, 32 mm in width, and 26 mm in depth. The blade head, the top part in the figure, is where fatigue cracks are more likely to occur. Figure 17 shows two Sonic IR images of the head section. They were taken from two different experiments: the left one with a non-chaotic excitation, and the right one with a chaotic excitation. In this sample, only three cracks can be detected with the non-chaotic excitation, which are circled in the image on the right as well. Not only did the three cracks show up with much lower signal level but also the inspection with non-chaotic sound missed many more cracks in other locations in the head, shown as bright spots. In such cases, the probability of detection (POD) of defects is largely reduced, which resulted in a high “miss” rate. Thus, chaotic excitation is a “must” to achieve a desirable POD.

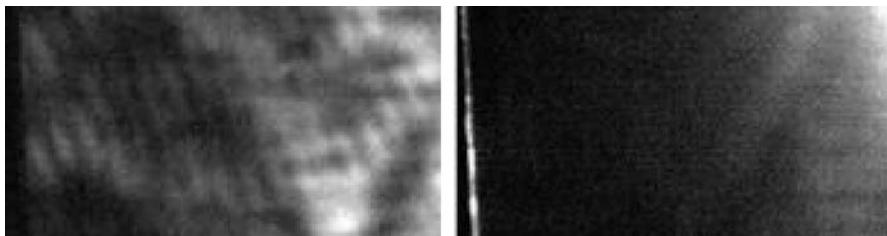
Chaotic excitation has another important advantage in the application of Sonic IR imaging evaluation of composite materials and/or regular geometry structures. For example, in laminated composite structures, part of the sound energy is absorbed. With non-chaotic sound, mode patterns are very often built up during the excitation pulse, as shown in the left image in Fig. 18. These mode patterns can easily mask the defect or part of the defect which results in no detection of the defect or lack of confidence in the call of the defect. On the other hand, chaotic excitation produces many more sounds with different frequencies. Because of that, mode patterns are

**Fig. 16** Photo of a commercial turbine blade for Sonic IR imaging evaluation





**Fig. 17** Two Sonic IR images of the blade head taken from two different experiments: the left image was taken with a non-chaotic excitation, whereas the right image was obtained with a chaotic excitation



**Fig. 18** Two Sonic IR images of a carbon fiber-reinforced composite panel from two different experiments: the left image, produced with non-chaotic excitation has mode patterns built up during the ultrasound pulse, while the right image, produced with chaotic excitation has much flatter background

usually suppressed, and thus produce much flatter background, as shown in the right image in Fig. 18. In such a case, defects such as delamination and/or disbonds are more likely to be detected.

---

## Summary

Sonic IR is a hybrid sensing technology that utilizes the concept of combining IR imaging with sonic/ultrasonic excitation. It is the youngest member in the NDE family. It can detect defects, such as cracks, delaminations, and disbonds in materials and structures including closed hairline cracks in critical materials and structures such as airplanes. The defects may be surface or subsurface. Sonic IR imaging has by far the broadest applications as well in comparison with the traditional NDE techniques since it can be applied to ferrous or non-ferrous materials, electrically conductive or electrically nonconductive materials, metal/metal alloy or nonmetal

materials, single materials or composite materials, regular geometry or nonregular geometry objects. The nonlinearity can be utilized for its advantage of enhancing the defect detection, and thus achieve high probability of detection. This technology is a wide-area, fast, and truly dark-field imaging method. Sonic IR imaging is by far the world's most sensitive and fastest crack detection technique.

---

## References

- Betz C (1997) Principle of magnetic particle testing, 2nd edn. Magnaflux Corporation
- Bray DE, Stanley RK (1997a) Part VI: Radiographic techniques in nondestructive evaluation. CRC Press, pp 429–507
- Bray DE, Stanley RK (1997b) Part II: Ultrasonic techniques in nondestructive evaluation. CRC Press, pp 51–169
- Bray DE, Stanley RK (1997c) Part V: Eddy current techniques in nondestructive evaluation. CRC Press, pp 367–427
- Bray DE, Stanley RK (1997d) Part IV: Magnetic flux leakage techniques in nondestructive evaluation. CRC Press, pp 215–363
- Bray DE, Stanley RK (1997e) Part VII: Penetrant techniques in nondestructive evaluation. CRC Press, pp 513–560
- Favro LD, Han X, Ouyang Z, Sun G, Sui H, Thomas RL (2000) Infrared imaging of defects heated by a sonic pulse. *Rev Sci Instrum* 71:2418
- Favro LD, Han X, Li L, Ouyang Z, Sun G, Richards A, Thomas RL (2001a) Review of progress in quantitative NDE vol 20, ed. by D.O. Thompson and D. Chimenti, CP557. Am Inst Phys 478–482
- Favro LD, Thomas RL, Han X, Ouyang Z, Newaz G, Gentile D (2001b) *Int J Fatigue* 23:S471
- Han X (2007) Sonic infrared imaging: a novel NDE Technology for Detection of crack/delamination/Disbond in materials and structures. In: Chen CH (ed) Ultrasonic and Advanced Methods in NDT. World Scientific Publishing
- Han X, Favro LD, Ouyang Z, Thomas RL (2001) Thermosonics: Detecting cracks and adhesion defects using ultrasonic excitation and infrared imaging. *J Adhes* 76:151
- Han X, Thomas RL, Favro LD (2002a) US Patent # 6,399,948, Miniature contactless sonic IR devices for remote non-destructive inspection
- Han X, Li W, Zeng Z, Favro LD, Thomas RL (2002b) Acoustic Chaos and sonic infrared imaging. *Appl Phys Lett* 81:3188
- Han X, Loggins V, Zeng Z, Favro LD, Thomas RL (2004a) Mechanical model for the generation of acoustic Chaos in sonic IR imaging. *Appl Phys Lett* 85:1332
- Han X, Zeng Z, Li W, Islam MS, Lu J, Loggins V, Yitaben E, Favro LD, Newaz G, Thomas RL (2004b) Acoustic Chaos for enhanced detectability of cracks by sonic infrared imaging. *J Appl Phys* 95:3792
- Han X, Islam MS, Newaz G, Favro LD, Thomas RL (2005) Finite-element modeling of acoustic Chaos in sonic infrared imaging. *J Appl Phys* 98:14907
- Han X, Islam MS, Newaz G, Favro LD, Thomas RL (2006) Finite-element modeling of the heating of cracks in sonic infrared imaging. *J Appl Phys* 99, 074905-1–074905-7
- Han X, Favro LD, Thomas RL (2010a) “Sonic IR imaging of delaminations and disbonds in composites”, invited paper. *J Phys D Appl Phys* 44:3
- Han X, He Q, Li W, Newaz GM, Favro LD, Thomas RL (2010b) Progress on developing sonic infrared imaging for defect detection in composite structures. In: AIP, Review of progress in quantitative nondestructive evaluation. AIP Press, Melville
- Han X, Zhao S, Zhang D, Lubowicki A, Favro L, Thomas R, Newaz G (2012) Develop sonic infrared imaging NDE for quantitative assessment on damage in aircraft composite structures. *Quan InfraRed Thermography* J

- Henneke EG II, Reifsneider KL, Stinchcomb WW (1979) Thermography- an NDI method for damage detection. *J Metals* 31:11
- Lovejoy D (1991) Penetrant testing – a practical guide. Chapman & Hall, London, p 53
- Metals Handbook (1989) Ultrasonic inspection. In: Nondestructive evaluation and quality control, vol 17, 9th edn. American Society for Metals, Metals Park, pp 231–277
- Mian A, Han X, Islam M, Newaz G (2004) Fatigue damage detection in graphite/epoxy composites using sonic infrared imaging technique. *J Compo Sci Technol* 64:657
- Mignogna RB, Green RE Jr, Duke JC Jr, Henneke EG II, Reifsneider KL (1981) Ultrasonics 19:159
- Phillips LC (1998) Eddy current methods in processing. In: Sensing for materials characterization, processing, and manufacturing. ASNT, Columbus
- Reifsneider KL, Henneke EG II, Stinchcomb WW (1980) In: Stinchcomb WW (ed) The mechanics of nondestructive testing. Plenum Press, New York, pp 249–276
- Song Y, Han X (2013a) The effect of loading force on ultrasound transducer non-linear coupling in sonic infrared imaging. *Int J Mod Eng* 14(1):88
- Song Y, Han X (2013b) Comprehensive study of the non-linear coupling effects between an ultrasound transducer and a target under inspection in sonic IR imaging. *Int J Eng Res Innov* 5(1):90
- Thomas RL, Favro LD, Han X, Ouyang Z, Sui H, Sun G (2001) US Patent # 6,236,049, Infrared imaging of ultrasonically excited subsurface defects in materials
- Wu D, Rantala J, Karpen W, Zenzinger G, Schoenbach B, Rippel W, Steegmueller R, Diener L, Busse G (1996) Review of progress. In: Quantitative nondestructive evaluation, eds., Donald O. Thompson and Dale E. Chimenti, vol 15, 511
- X-ray image was provided by Mr. Dan Fahrer



# Induction Thermography of Surface Defects

41

Udo Netzelmann

## Contents

Introduction to Induction Thermography .....	1498
Electromagnetic and Thermal Key Parameters .....	1499
Thermal Response to Inductive Heating in Various Materials .....	1500
Instrumentation .....	1504
Inductors .....	1505
Infrared Cameras .....	1506
Pre-Processing of the Recorded Image Sequences .....	1506
Defect Detection .....	1507
Detectable Types of Defects .....	1507
Detectable Defect Sizes .....	1508
Determination of the Crack Depth .....	1509
The Role of the Crack Width .....	1510
Crack Detection under Nonconducting Coatings .....	1511
Typical Applications .....	1511
Steel Components .....	1511
Detection of Hidden Defects in Ferritic Steel .....	1513
Testing in Relative Movement and HandHheld Induction Systems .....	1513
Semiconductors .....	1515
Carbon Fiber-Reinforced Polymer (CFRP) .....	1515
Railway Component Testing .....	1516
Comparison of Induction Thermography and Magnetic Particle Testing .....	1519
Standardization .....	1520
Summary and Outlook .....	1520
References .....	1520

---

U. Netzelmann (✉)

Fraunhofer-Institute for Nondestructive Testing IZFP, Saarbrücken, Germany

e-mail: [udo.netzelmann@izfp.fraunhofer.de](mailto:udo.netzelmann@izfp.fraunhofer.de)

**Abstract**

A survey on theory, characteristic quantities, and the experimental technique of induction thermography is given. Induction thermography is used for surface defect detection in forged parts of ferromagnetic steel at typical frequencies of 100–300 kHz. Values for the detection limits for various types of cracks and approaches to determine crack depths are given. The sensitivity for crack detection is comparable to magnetic particle inspection. A hidden defect in ferritic steel with a coverage of 140 µm was detected by lowering the induction frequency down to 1500 Hz. Cracks in silicon solar cells were detected. Defects of fibers were detected in carbon fiber-reinforced polymer (CFRP). Inductive excitation is complementary to flash excitation. Crack detection in railway components like rails and wheels is shown. In rails, a larger defect could be detected from a test car moving at a speed of up to 15 km/h. A fully automated demonstrator for wheel testing was built up, which can detect surface defects in railway wheels with sensitivity comparable to magnetic particle testing. Standardization of thermography has gained progress in the last years and led to first standards on active thermography and induction thermography.

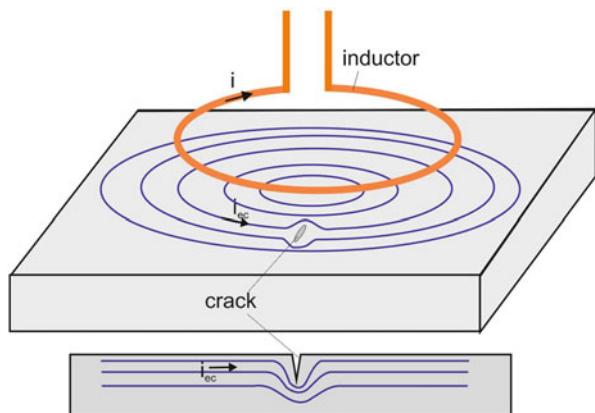
---

**Introduction to Induction Thermography**

Induction thermography or pulsed eddy current thermography uses electromagnetic pulses to excite eddy currents in electrically conductive materials. An inductor (Fig. 1) is positioned at a certain distance from a surface to be tested for surface defects like cracks. In the test object, eddy currents  $i_{ec}$  are induced that form closed loops in the material. The currents have to circumvent the cracks. This causes local changes of the electrical current densities in the material. The eddy currents generate heat by resistive losses. The heat can be detected by an infrared camera observing the surface. Around cracks, the current density is changed, e.g., increased at the crack tips. As heat generation is proportional to the square of the current density, the cracks will form a characteristic thermal pattern on the surface which has a specific time dependence that reveals the position and the orientation of the crack.

First applications in the steel industry were reported more than three decades ago (Kremer 1984). The Therm-O-Matic system was applied to detect longitudinal cracks by continuous inductive heating of moving steel bars. The speed relative to the coil was up to 1 m/s. Several of these systems are still in operation today. Further, the technique has been applied using periodic heating and phase sensitive detection for characterization of the adhesion of rubber on aluminum (Heath and Winfree 1990). A ferrite core transducer was used for excitation. Pulsed high-frequency burst excitation is used in the production process for detection of cracks in turbine blades (Bamberg et al. 1999). These tests are part of the internal component qualification by the manufacturer. Further applications were reported on compressor blades and tensile test specimens, as well as on impact damage detection in carbon fiber reinforced polymers (Riegert et al. 2004). There was significant work on analytical

**Fig. 1** Scheme of induction thermography



and numerical modelling of the signal from cracks in steel (Oswald-Tranta 2004). The current distribution and the resulting heating were calculated, providing insight into the contrast formation. The signal contrast was analyzed experimentally as a function of crack length, crack depth, and crack orientation (Walle and Netzelmann 2006). In Vrana et al. (2008), the current flow in small components was studied including the proximity effect. The contrast from cracks filled with conducting materials was investigated. Recent applications were devoted to testing of adhesive bonds of aluminum sheets for automotive application (Balaji et al. 2013) and describe carbon fiber laminate testing (He et al. 2014).

In this contribution, first, the characteristic electric and thermal quantities will be discussed. After a description of the experimental setup, detection of various types of cracks will be described. Then examples will be shown for different application fields of induction thermography. A final remark is given to the present efforts for standardization.

## Electromagnetic and Thermal Key Parameters

In contrast to eddy current testing, induction thermography has both an electromagnetic and a thermal aspect (Netzelmann and Walle 2008). In order to estimate the ability to detect defects in materials by induction thermography, one has to consider the electromagnetic skin depth  $\delta$ , which describes the depth of the induction current flow. For a thick plane body, the electric and magnetic fields and the resulting currents are exponentially damped with the skin depth as a characteristic length, when propagating into the depth. It is given by

$$\delta = \sqrt{\frac{1}{\pi f \sigma_e \mu}}, \quad (1)$$

where  $\sigma_e$  is the electrical conductivity,  $\mu$  the magnetic permeability, and  $f$  is the electromagnetic frequency. Another characteristic length is the thermal penetration depth  $\mu_T$ . The depth of the thermal propagation for a given observation time  $t$  (in a pulsed experiment) is given by

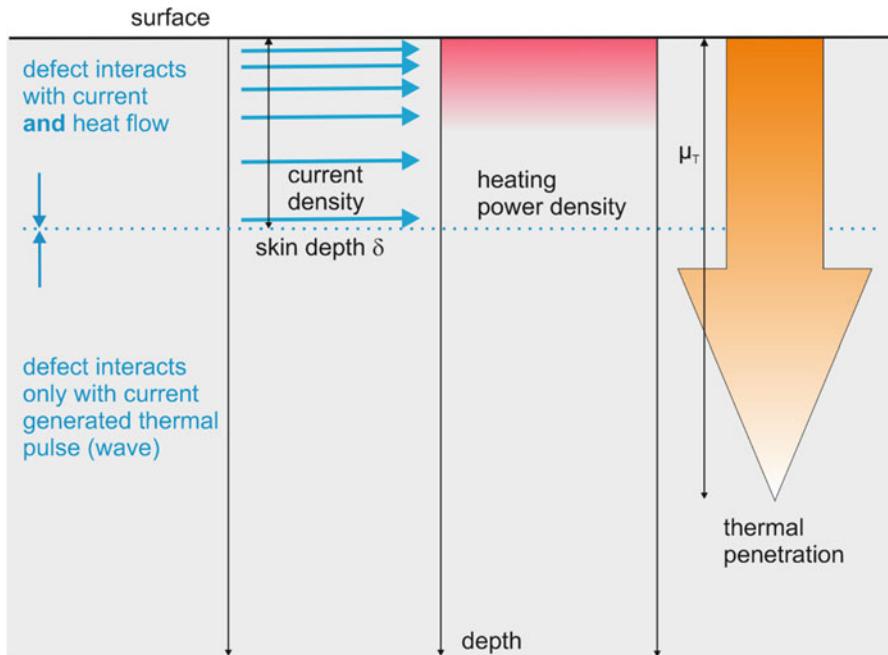
$$\mu_T(t) = 2\sqrt{\alpha t} \quad (2)$$

Here,  $\alpha$  is the thermal diffusivity of the material. In Fig. 2, the distribution of the induction currents, the heat sources, and the resulting thermal flow are sketched.

Examples for skin depth and thermal penetration depth for different materials for an induction frequency of  $f = 100$  kHz and an observation time  $t = 100$  ms are given in Table 1.

### Thermal Response to Inductive Heating in Various Materials

Basic requirement for application of induction thermography is that there is an electrically conducting material. When a plane electromagnetic wave enters a conducting half-space, its field components and the induction currents are damped with the skin depth  $\delta$  as characteristic length. The heat sources due to ohmic losses are proportional to the square of the induced currents. Therefore, the heat source



**Fig. 2** Scheme of inductive heating and thermal propagation in a conducting material

**Table 1** Material parameters, characteristic lengths and induction heating efficiency

Material group		Electrical conductivity in $10^3 \text{ S/m}$	Relative magnetic permeability	Thermal conductivity in $\text{W/(mK)}$	Electro-magnetic skin depth in mm (100 kHz)	Thermal penetration depth in mm ( $t = 0.1 \text{ s}$ )	Surface temperature rise in K after 0.1 s heating with 0.02 T induction field amplitude at 100 kHz
<b>I</b>	<b>Cast iron</b>	6.2	200	49	0.045	2.43	12.4
	<b>Nickel, pure</b>	14.62	100	90.7	0.042	3.03	3.9
<b>II</b>	<b>Silver, pure</b>	62.87	1	408	0.201	8.15	0.1
	<b>Zinc, rolled</b>	16.24	1	113	0.395	4.06	0.4
	<b>Aluminium</b>	22.53	1	177	0.335	5.97	0.3
	<b>2014-T6</b>						
	<b>Copper, pure</b>	60.09	1	401	0.205	6.83	0.1
<b>III</b>	<b>Inconel 600</b>	0.98	1	14.9	1.608	1.26	1.6
	<b>Stainless steel 316</b>	1.33	1	13.4	1.380	1.18	1.5
	<b>Titanium 6AL-4 V</b>	0.58	1	7.2	2.090	1.08	2.9
<b>IV</b>	<b>Carbon fibre reinforced polymer perp.</b>	0.0001	1	0.7	159	0.39	5.3
	<b>To fiber</b>						
	<b>SiC ceramic</b>	0.00005	1	80	225	3.36	3.5
	<b>Silicon</b>	0.001	1	148	50	6.11	5.8

distribution decays with a damping constant of  $\delta/2$  into the depth  $z$  of the material. The heat source  $q(z)$  can be described by

$$q(z) = \frac{b^2}{2\mu_0^2} \mu_0 \mu_R 2\pi f \exp\left(-\frac{2}{\delta} z\right). \quad (3)$$

Here,  $b$  is the high-frequency magnetic field at the sample surface,  $\mu_0$  the permeability constant, and  $\mu_R$  the high-frequency magnetic permeability ( $\mu_R = (\mu_r'^2 + \mu_r''^2)^{0.5} - \mu_r''$ , where  $\mu_r'$  and  $\mu_r''$  are the real and the imaginary part of the complex relative permeability, respectively). At typical induction frequencies,  $\mu_R$  is usually somewhat smaller than the static relative permeability (Bowler 2006).

When induction heating starts at the time  $t = 0$ , the resulting surface temperature change  $\Delta T$  at the surface can be calculated based on a result in Carslaw and Jaeger (1959):

$$\Delta T = \frac{b^2}{2\mu_0^2} \mu_0 \mu_R \frac{2\pi f}{\lambda} \delta^2 \left( \frac{1}{2\sqrt{\pi}} \frac{\mu_T}{\delta} - \frac{1}{4} \left( 1 - \exp\left(\left(\frac{\mu_T}{\delta}\right)^2\right) \operatorname{erfc}\left(\frac{\mu_T}{\delta}\right) \right) \right). \quad (4)$$

Here,  $\lambda$  is the thermal conductivity of the material. The time dependence of  $\Delta T$  is implicitly given by  $\mu_T$  (Eq. 2). The temperature change is governed by the ratio of thermal penetration depth and electromagnetic skin depth. Equation 4 can be used to compare the thermal response of a material after a certain heating time caused by a given induction field at its surface (in the following called “induction heating efficiency”). Table 1 shows examples for material parameters and penetration depths and identifies several material groups in induction thermography. In its right column, Table 1 lists the surface temperature rise  $\Delta T$  resulting from an induction field of  $b = 0.02$  T at an induction frequency of  $f = 100$  kHz after 0.1 s of inductive heating.

Four groups of materials can be identified in Table 1. Due to their high magnetic permeability, the ferromagnetic metals (group I) even after a short time have a very small skin depth compared to the thermal penetration. Only a thin surface layer generates the heat. The nonmagnetic metals in group II are often good electrical conductors, but their relative permeability is close to one. Their skin depth is still one order below the typical thermal penetration. In group III there are metallic alloys with relatively poor electrical and thermal conduction. Here, the electromagnetic and thermal penetration depth is of comparable size for the given parameters. Finally, the materials of group IV have an electromagnetic skin depth significantly larger than the thermal penetration.

The right column in Table 1 shows that the ferromagnetic metals form group I have the highest induction heating efficiency. On the other hand, the induction heating efficiency in group II is poor. In particular the aluminum alloys have efficiency about a factor of 30 below that of magnetic steel, rendering testing of these materials difficult. This is further complicated by their often low infrared emissivity. In spite of this, successful applications on aluminum were reported (Tsopelas and Siakavellas 2011).

For the materials in groups I and II after a very short time the thermal penetration is larger than the skin depth, ( $\mu_T > > \delta$ ). Then, Eq. 4 can be simplified to

$$\Delta T(t) = \frac{b^2}{2\mu_0^2} \mu_0 \mu_R 2\sqrt{\pi f} \delta \frac{\sqrt{t}}{\sqrt{(\lambda \rho c)}}. \quad (5)$$

Here,  $\rho$  and  $c$  are the density and the specific heat of the material, respectively. Experimentally, the surface heating process is characterized by a temperature rise proportional to the square root of time. As the skin depth  $\delta$  is proportional to  $f^{-0.5}$ , the temperature rise  $\Delta T$  is in total proportion to the square root of the induction frequency.

Due to their larger skin depth, the alloys from group III represent a transition from surface heating to volume heating. The induction heating efficiency is much better than that of group II. Indeed, many successful applications of crack detection in aeronautical or power generation turbine components with materials from group III were reported (Zenzinger et al. 2007).

Finally, the materials in group IV are characterized by  $\mu_T << \delta$ , even for longer times  $t$ . In this case, Eq. 4 can be simplified to

$$\Delta T(t) = \frac{b^2}{2\mu_0^2} \mu_0 \mu_R 2\pi f \frac{t}{\rho c}. \quad (6)$$

The surface temperature rises proportional to time, which is typical for a volume heat source. It is proportional also to the induction frequency. The induction heating efficiency does not reach that of magnetic steel but is better than that of the materials in group III. Successful applications, e.g., in carbon fiber-reinforced polymers, were reported (Liang et al. 2016; Guo et al. 2017).

The total power density  $S$  in W/m<sup>2</sup> absorbed by the sample can be obtained by integration of Eq. 3 over the depth  $z$ . It is given by

$$S = \frac{b^2}{2\mu_0^2} \sqrt{\frac{\pi \mu_0 \mu_R f}{\sigma_e}} \quad (7)$$

In all cases, higher magnetic permeability increases the signal, which explains the good induction efficiency in ferritic steel. For magnetic materials, Eqs. 3, 4, 5, 6, 7 are an approximation for small deviations of the magnetization from its equilibrium. It has been shown that static magnetic fields applied in addition to the high-frequency field can improve the thermographic contrast of cracks, when applied in a proper direction (Jäckel and Netzelmann 2013).

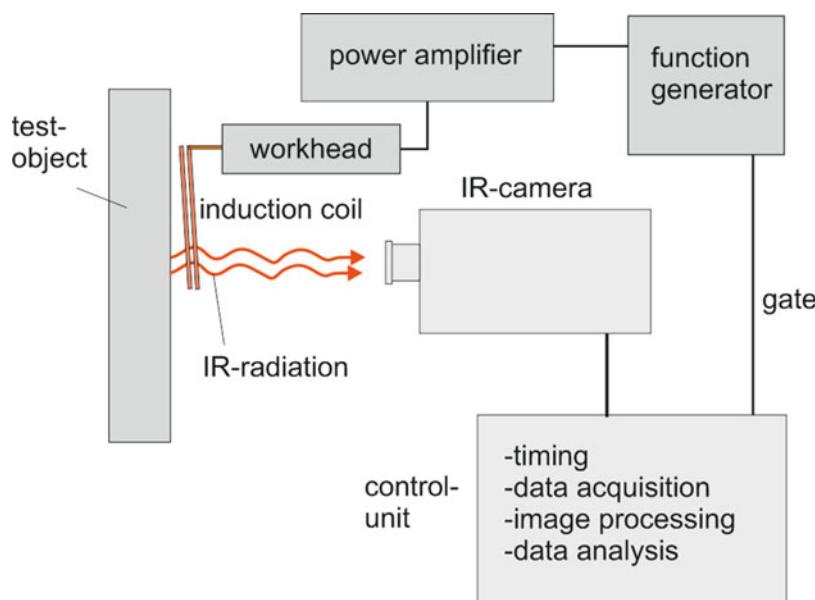
It should be mentioned that the electromagnetic skin depth as defined in Eq. 1 is valid for a plane electromagnetic wave incident to a plane, thick conductor. The damping lengths will be somewhat different for rod and tube geometries, for thin plates, small components, and under inductor wires. Even more complicated are the current paths in anisotropic conductors like CFRP.

## Instrumentation

A typical experimental setup with a pulsed excitation system is shown in Fig. 3. A high frequency induction coil in the neighborhood of the test object generates eddy currents in burst pulses of typically 50–500 ms length. The induction coil and capacitive elements in the workhead form a resonant circuit with large rf currents close to resonance. Pulsed excitation is the fastest way for testing of components, in particular, when mass products have to be inspected in the industrial environment. It should be noted that the resonant circuit consisting of workhead and coil needs some minimum settling time between zero and full induction power and vice versa. This may limit detection of smallest defects. The burst pulse length should be short enough not to miss the smallest defects but long enough to generate sufficient defect contrast.

Alternatively, excitation can be generated by nonresonant high power amplifiers (preferably in the frequency range  $< 50$  kHz). A fast rise-time of the induction current is possible.

Another possibility, in particular when the available maximum induction power is low, is amplitude modulation of the high-frequency signal at a lower modulation frequency  $f_{\text{mod}}$ . A gateable frequency generator can be used, followed by a broadband power amplifier. Then, a lock-in algorithm is applied to the recorded time-depended signal of each pixel in order to calculate amplitude and a phase image (Riegert et al. 2004). Periodic excitation is used mainly in the laboratory or at induction frequencies, where high-power pulses are hardly available. Compared to



**Fig. 3** Scheme of a set-up for induction thermography

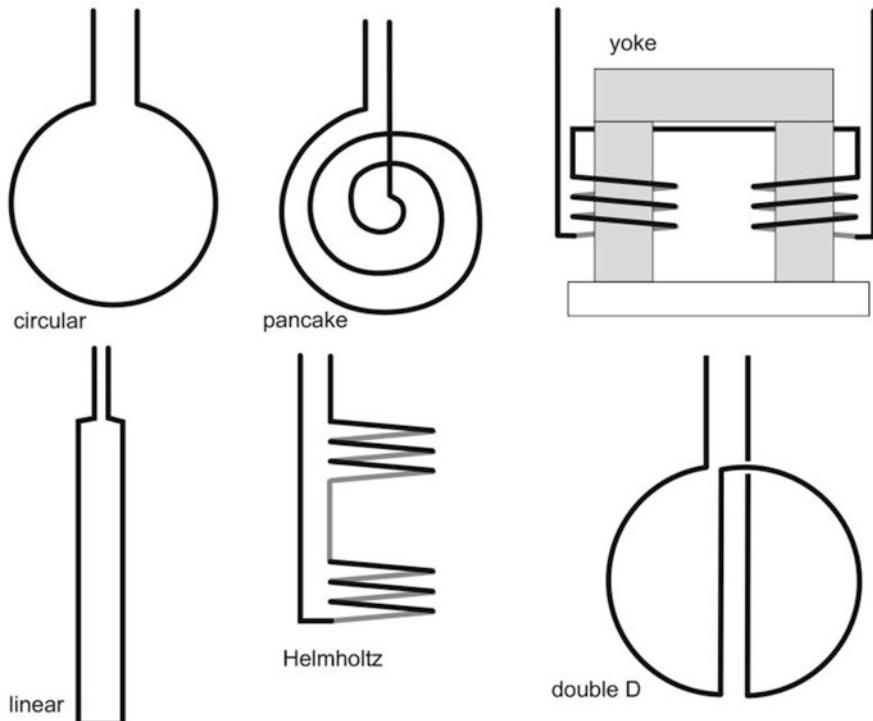
pulsed operation, the measurement time to achieve a given signal/noise ratio will be generally longer.

Camera, induction coil, and test object can also be in relative movement. Then, the induction coil is operated in a cw mode. For testing of long tubes and steel bars, encircling coils are used. A time dependent excitation is caused by the relative movement (see a later section).

A less common possibility to generate time-dependent heat flows is to move a permanent magnet with sufficient relative speed to the test object. This will induce eddy currents that generate heat (principle of the eddy-current brake) (Netzelmann 2006).

## Inductors

The induction coils used can be similar to those used for induction hardening or other industrial heating tasks (Fig. 4). They can be air coils with or without water cooling. Water cooling will be necessary for continuous operation of the coil, e.g., for relative movement between test object and coil, or for fast testing cycles. Helmholtz coils were reported that have the advantage to homogenize the excitation field and to allow full optical access to the component surface for the infrared camera



**Fig. 4** Typical inductors as used for induction thermography

(Wilson et al. 2010). The magnetic flux of a coil can be concentrated by using a yoke made out of a material with high permeability. Linear coils can be used for inspection objects on a moving conveyor belt. Double D coils homogenize the magnetic field in the area between the parallel wires. Other inductor concepts use two orthogonally oriented yokes that are fed by currents which are 90° phase shifted in order to form a rotating induction field (Tang et al. 2018). Using this, the crack signal is less sensitive to the crack orientation versus the induction current.

Covering the winding of the coil by a thermally thick insulator helps reduce time-dependent thermal radiation from resistive losses in the coil.

## Infrared Cameras

The infrared camera is used to record image sequences of the surface radiation from shortly before until sometime after the heating pulse (Fig. 3). Two types of detectors are common:

The bolometer detector is a thermal detector for infrared radiation and allows frame-rates up to about 60 Hz. They are typically sensitive in the 7–15  $\mu\text{m}$  wavelength range. These detectors have seen significant technical progress with respect to sensitivity and pixel resolution. Models with  $1920 \times 1080$  pixels resolution have become available. The noise equivalent temperature deviation (NETD) may go down to 30 mK or less. Thermal detectors represent the most cost-efficient and a low-maintenance solution. They are sufficient to solve a large class of testing applications.

Infrared cameras based on photon detectors are more sensitive (NETD down to 10 mK), show less or no crosstalk between neighboring pixels, and offer short integration times and high frame-rates. They are typically sensitive in the 2–5.5  $\mu\text{m}$  wavelength range (mid-wave infrared, MWIR) but also available for the longer wavelengths. A cooling system is required. The resolution currently goes up to  $640 \times 512$  pixels or higher. For detection of smallest defects, this camera type is the best choice. The price is significantly higher than that of a bolometer camera and the detector coolers have limited lifetime.

Both camera types need a possibility for synchronization of the image acquisition (at least by a frame time-stamp) with the excitation. Bolometer cameras must allow the deactivation of the built-in shutter for drift- and nonuniformity correction during the time of image acquisition.

For induction thermography, photon detector cameras will be preferably used for testing materials with high thermal conductivity and low induction efficiency, whereas bolometer cameras may be sufficient for materials with high induction efficiency or for poor conductors like CFRP.

## Pre-Processing of the Recorded Image Sequences

A pre-processing of the recorded image sequences has as a major task to suppress the local inhomogeneity of inductive heating and to extract time-dependent signals that

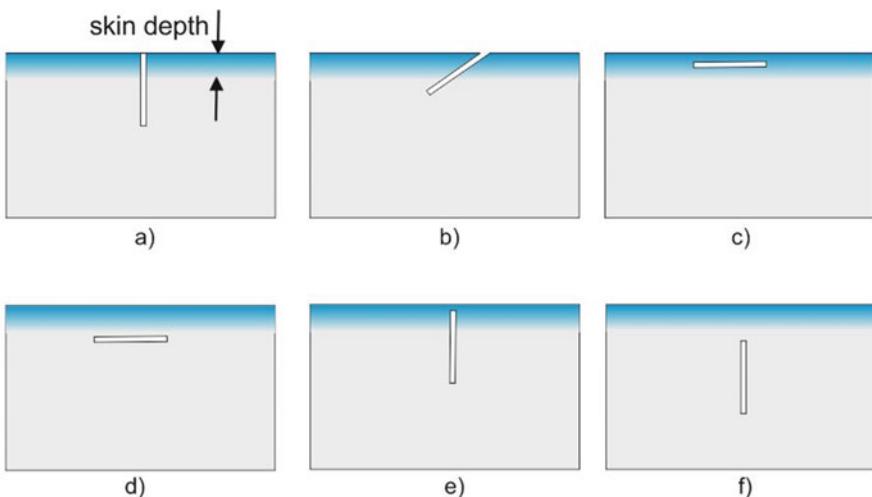
are characteristic for many defects. A frequently used technique applied in the case of excitation by high-frequency bursts is the pulse-phase transform (PPT). The time dependent signal from each pixel in the image is transformed into the frequency space. From the results, amplitude and phase images in the frequency space are generated. A common practice is to transform a time interval of  $t_p + t_c$ , where  $t_p$  is the length of the burst pulse and  $t_c$  a cooling time after the pulse. Amplitude and phase images at the frequency of  $1/(t_p + t_c)$  and its multiples will result. The phase images will largely be free of the inhomogeneous distribution of induction currents (which is almost unavoidable in induction thermography) or of local variations of surface emissivity. Characteristic phase shifts around cracks will appear more clearly with better signal/noise ratio. Varieties of other pre-processing techniques like principal component analysis are available and have been investigated.

In the case of periodic modulation, conventional processing of the recorded image sequences by lock-in algorithms is performed.

## Defect Detection

### Detectable Types of Defects

In a previous chapter, it has been pointed out that the interaction with defects in induction thermography is based on electromagnetic and thermal interaction with the defect (Fig. 2). These interactions determine the detectability of defects. In Fig. 5, typical defects are shown in a cross-section through an electrically conducting material. The skin depth  $\delta$  is indicated as a zone at the surface.



**Fig. 5** Some types of cracks to be detected by induction thermography. Types (a) to (e) are basically detectable, type (f) is not directly detectable

The perpendicular crack (a) interacts with the induced current flow and then with lateral heat flows and can therefore be detected. The slanted crack (b) interacts with the current flow, but also with the heat, which is captured in the wedge between crack and surface. Slanted cracks are even better detectable than cracks perpendicular to the surface, which favors detection of forging or rolling laps. Crack type (c) would hardly be detectable by eddy current techniques, but it will produce a thermal contrast, as the heat generated between crack and surface cannot directly flow into the depth. The crack is detected only thermally, comparable to optical surface heating. This is also true for case (d), as long as the crack is located within the thermal diffusion length and its lateral dimension is larger than the coverage. The hidden crack (e) will be detectable as long as the coverage is smaller than half the skin depth (Walle et al. 2012). In the case (f), the crack has neither electromagnetic nor thermal interaction and will therefore not generate a thermal contrast. But case (f) can be converted to case (e), when the skin depth is increased by lowering the induction frequency. An example will be given in the application section.

Another aspect is the orientation of the induced currents with respect to the crack direction. Generally, the thermal contrast will be weaker if the induction current is incident in the crack direction and it will be maximal if it is incident perpendicular to the crack. For ferritic steel, several studies showed that if the angle between crack direction and direction of the induced currents is varied from  $90^\circ$  to  $0^\circ$ , the thermal contrast decreases from 100% down to about 30%, which still allows crack detection with reduced signal/noise ratio (Walle and Netzelmann 2006). Nevertheless, for highest detection sensitivity, it may be necessary to perform two subsequent measurements with orthogonal field orientation or to use an inductor with a rotating induction field.

## Detectable Defect Sizes

Validation studies were performed with the aim to determine detection limits of induction thermography. The materials studied were ferritic steel S235, austenitic steel X5CrNi18–10, copper Cu-ETP F25, and aluminum AlNg3. Test bodies from these materials were manufactured including artificial cracks (short notches with a width of 0.3 mm) produced by spark eroding. Perpendicular and slanted cracks of different depths and lengths were investigated. From the measurements, the S/N ratio achieved was determined.

The S/N is usually determined from  $C_a/\Delta T_n$ , where  $\Delta T_n$  is the noise level and  $C_a$  is the absolute defect contrast

$$C_a = \Delta T_{\text{def}} - \Delta T_s. \quad (8)$$

Here,  $\Delta T_{\text{def}}$  is the temperature change at the defect position and  $\Delta T_s$  is the temperature change without the defect. The latter can approximately be determined in a sound region close to the crack, which is not yet affected by the crack contrast. In an analog way, an absolute phase contrast of the defect can be determined.

Results for the S/N from artificial cracks in ferritic and austenitic steel obtained from the absolute temperature contrasts are summarized in Fig. 6 (Walle et al. 2009). The results show that the smallest cracks are detectable in ferritic steel and that inclined cracks are detected with more sensitivity than perpendicular cracks. The sensitivity of detection is comparable to that of MT.

For forged ferritic steel, it was shown that for small surface cracks, the signal contrast varies roughly with the product of crack length and crack depth, with saturation occurring towards very long and very deep cracks. Detection of very shallow and small cracks will require short induction pulses and fast camera frame rates.

## Determination of the Crack Depth

A major advantage of induction thermography vs. magnetic particle testing (MT) is due to the fact, that induction thermography has the potential to give information on the crack depth. This is hardly possible in MT. If this information were available, in many production processes it could be decided if a component with a detected surface crack is worth to be reworked by grinding a small layer of material. By this, the crack could be removed completely. While the focus in research on induction thermography has been the detection of defects, in the last years there was activity aiming at defect characterization, in particular depth determination of cracks.

<b>• surface cracks perpendicular to the surface</b>				
steel	crack depth [mm]	crack length L [mm]	S/N [dB]	
S235	$d \geq 0.15$	$L \geq 7.5$	> 10	
	$d \geq 0.25$	$2 \leq L < 7.5$	> 10	
X5CrNi18-10	$d \geq 0.3$	$L \geq 5$	6	
	$d \geq 0.5$	$2 \leq L < 5$	6	
<b>• slanted surface cracks (depth d measured perpendicular to the surface)</b>				
steel	crack depth [mm]	crack length L [mm]	inclination angle $\alpha$ [°]	S/N [dB]
S235	$d \geq 0.08$	$L \geq 7.5$	$20^\circ \leq \alpha \leq 60^\circ$	> 10
	$d \geq 0.15$	$2 \leq L < 7.5$	$20^\circ \leq \alpha \leq 60^\circ$	> 10
X5CrNi18-10	$d \geq 0.20$	$L \geq 5$	$30^\circ \leq \alpha \leq 60^\circ$	6
	$d \geq 0.35$	$2 \leq L < 5$	$30^\circ \leq \alpha \leq 60^\circ$	6
<b>• covered, perpendicular cracks</b>				
steel	cover: z	crack length [mm]	S/N [dB]	
S235	$z \leq 0.05$ mm	$L \geq 7.5$ mm	6	
X5CrNi18-10	$z \leq 1$ mm	$L \geq 7.5$ mm	6	

**Fig. 6** Experimentally determined detection limits in two types of steel for different types of cracks

The signal of a crack indication is usually characterized by the relative crack contrast

$$C_r = \frac{\Delta T_{\text{crack}} - \Delta T_s}{\Delta T_s}, \quad (9)$$

where  $\Delta T_{\text{crack}}$  is the temperature change at the crack. An analogous procedure is also common when phase images are evaluated.

A detailed analysis on crack depth determination in induction thermography can be found in Oswald-Tranta (2018). Infinitely long cracks were considered. Some results are cited in the following.

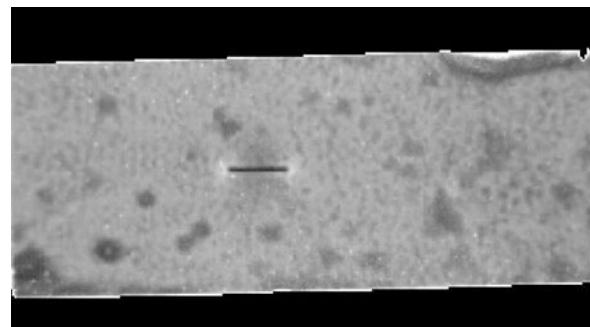
With the exception of a special case, there is no analytical formula for the crack contrast as a function of the crack depth  $d$ . Numerical simulations that were confirmed by experiments have the following results:

- The relative crack contrast  $C_r$  is between  $-1$  and  $+1$ .
- The crack contrast is determined by  $\delta/\mu_I$ , the ratio of crack depth  $d$  to the thermal diffusion length, the relation of the crack depth to the skin depth  $d/\delta$  and the heating pulse duration  $t_p$ .
- For ferritic steel (material group I in Table 1) and cracks with  $d > > \delta$ , the absolute phase contrast across the crack shows a maximum at the crack position and two minima beside the crack. The depth of the minima increases with crack depth  $d$ . There is a saturation effect for deeper cracks, which can be overcome by increasing the pulse length  $t_p$ , at the cost of probably missing shallow cracks.
- For nonmagnetic materials (material groups II and III in Table 1) and cracks with  $d \approx \delta$ , the absolute phase contrast shows a minimum at the crack position. The minimum is deeper with increasing crack depth  $d$ . Shorter pulse durations  $t_p$  increase the phase contrast.
- The relative amplitude contrast  $C_r$  may change its sign with time for larger skin depth. The absolute phase contrast may change its sign with increasing pulse duration  $t_p$  for larger skin depth.
- Slanted cracks show asymmetric contrast profiles and need more effort to extract the defect depth from the images.

## The Role of the Crack Width

There is a difference between artificial cracks that are machined into a surface and real cracks, which are more irregular in shape and sometimes have local contact points between the crack flanks. In thermography, an open natural crack and most artificial crack represent a line of higher emissivity in the thermography image. This is superimposed onto the contrasts caused by the disturbed current flow. This contrast decreases, when the crack width becomes smaller. On the other hand, the crack contrast can show a chain of local maxima along the crack, if there are several contact points between the crack flanks (Vrana et al. 2008). The induction currents

**Fig. 7** Thermographic phase image at an analysis frequency of 2.5 Hz. The images sample size is 40 mm × 20 mm



concentrate at these points and generate strong indication in the thermal image. This was also observed for open cracks, which were contaminated by metal chips and dust.

### Crack Detection under Nonconducting Coatings

Often surface defects in metals have to be detected under nonconducting coatings. Examples are cracks in turbine blades with ceramic thermal barrier coatings or corrosion spots under paint layers. Defects under coatings can be detected well as long as the coating is thermally thin. Then, the crack is even detected with higher S/N due to the increased surface emissivity (the well-known effect of surface blackening). Even for thermally thick coatings there is a chance for detection, if the coating is partially transparent for the thermal infrared. This is the case for ceramic thermal barrier coatings in the MWIR. But also polymer coatings are often partially transparent in the thermal infrared (Wang et al. 2018). Figure 7 shows the signal of an artificial crack with 5 mm length, 3 mm depth, and 0.5 mm width in ferritic steel obtained by pulse-phase thermography with  $t_p = t_c = 0.2$  s. The crack is hidden under a polymer coating of 0.36 mm thickness.

The weak structures in the area around the crack are due to a lack of adhesion of the coating on the steel.

---

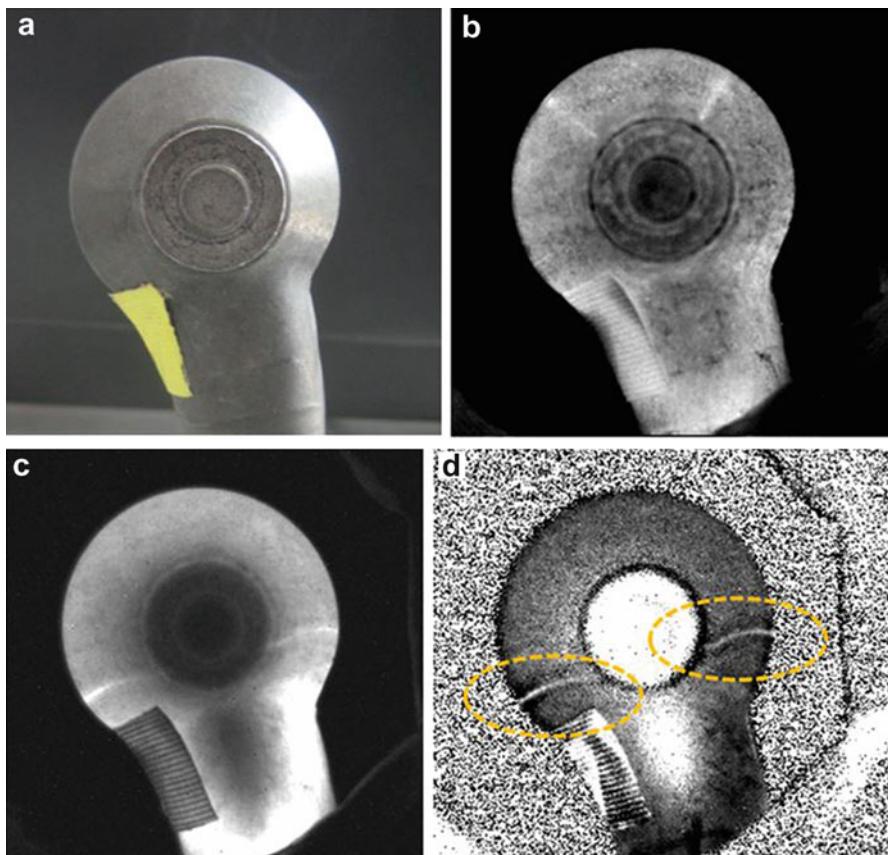
## Typical Applications

### Steel Components

A study on forged steel parts of very different shapes (Walle and Netzelmann 2006) has shown that typical cracks occurring in such parts can be detected well by induction thermography. There was a first reliability study that indicated a good detection rate and low false alarm rate for induction thermography, whereas MT, as performed under production conditions, often suffered from a high number of false alarms.

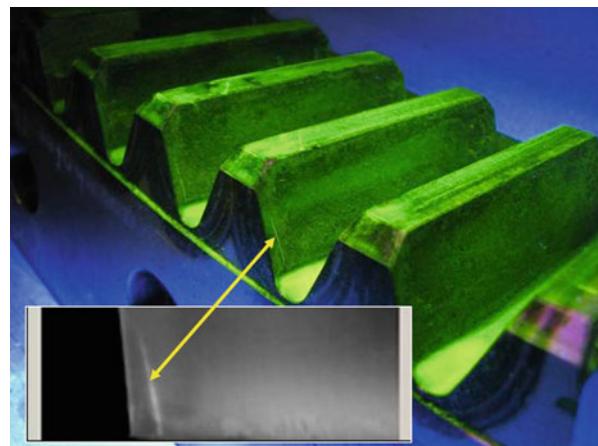
In this and other studies, mainly pulsed high-frequency bursts were used, which are probably the best approach, when the available testing time is small. Using periodically modulated high-frequency signals and lock-in processing, crack testing can be achieved at much lower excitation powers of some tens of Watts or less (Riegert et al. 2004). The example shown in Fig. 8 was obtained at about 20 W induction power with a measurement time of 3.5 s. It can be seen that two radially oriented cracks are detected with good contrast in the amplitude and the phase image (Fig. 8c, d). Contrasts on the conical surface (Fig. 8b) due to reflection of external heat sources (in this case lamps in the laboratory) are suppressed by the lock-in calculation.

Another application demonstrates the detection of hardening cracks on a large gear-tooth (Fig. 9). The crack indication obtained by pulsed induction thermography is shown along with the conventional MT image as obtained under UV light. A concept for an automated manipulator was developed that scans all relevant faces of the test object.



**Fig. 8** (a) Photo of the forged part. (b) Thermographic image before inductive heating, (c) Thermographic amplitude image at 4.5 Hz, (d) Thermographic phase image at 4.5 Hz. Image size 57 mm × 60 mm. Crack locations are marked

**Fig. 9** Hardening crack on a gear-tooth detected by induction thermography (small inset) and by magnetic particle testing (large picture). Image size (inset): 22 mm × 60 mm

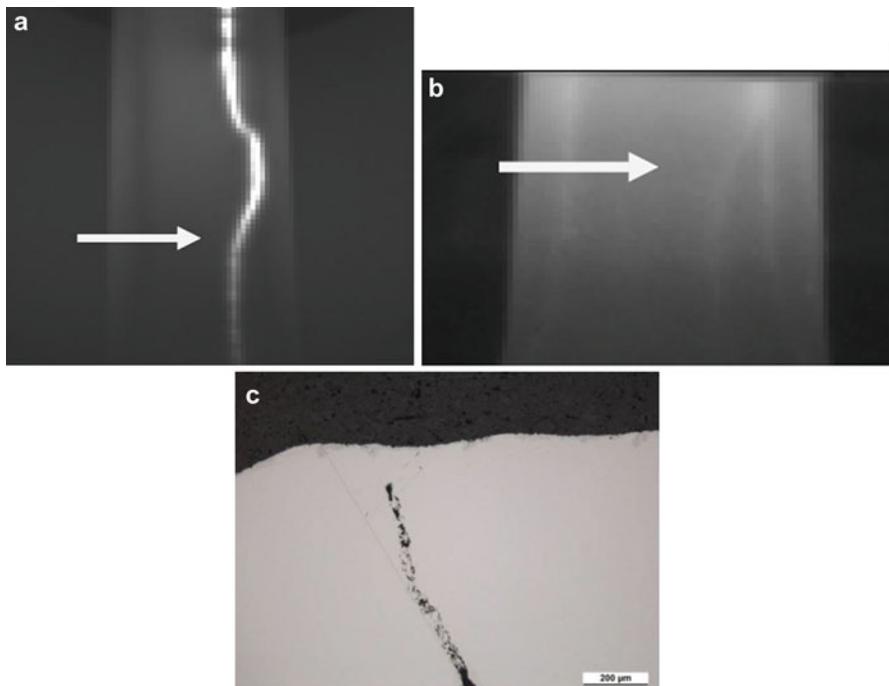


### Detection of Hidden Defects in Ferritic Steel

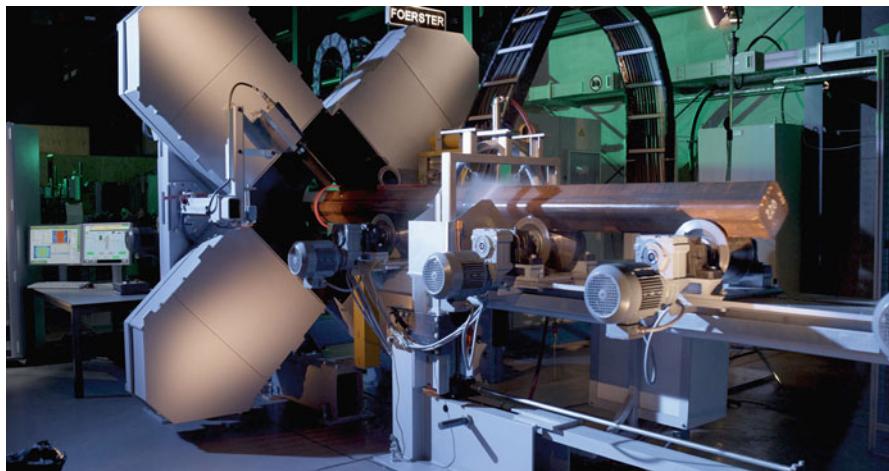
As discussed before, for hidden cracks perpendicular to the surface, the only way to detect the crack by induction thermography is to enlarge the skin depth to at least half of the crack coverage. As a typical skin depth in ferritic steel is about 50 µm, low-frequency power induction generators are required to meet this condition for cracks with coverage of some 100 µm. Figure 10 shows thermographic images of a natural hidden crack with a distance to the surface of 140 µm (as later verified by metallography, Fig. 10c). At 100 kHz induction frequency, there is only a weak contrast (Fig. 10b). If the induction frequency is reduced to 1500 Hz, there is sufficient electromagnetic interaction with the crack and a large crack contrast is obtained. The results were further supported by numerical simulations (Walle et al. 2012). Frequencies down to 300 Hz were used to detect hidden notches. A ratio of coverage to skin depth smaller than 0.5 is required to detect a hidden vertical crack. In the case shown in Fig. 4a, this ratio was 0.34. For the low-frequency experiments, a high-power audio-frequency amplifier was employed.

### Testing in Relative Movement and HandHeld Induction Systems

Following the success of the Them-o-Matic (or ELKEM) system (Kremer 1984), several systems with modern infrared cameras and advanced processing were set up. One example is the system DEFECTOVISION® IR by Institut Dr. Foerster GmbH & Co. KG, Germany (Fig. 11). Long round and square steel bars products are inspected for surface defects. Not only single thermographic images but sequences of images are processed. This allows suppression of false indications and an estimation of the defect depth. Bar dimensions of 50 to 220 mm can be tested. The testing speed is up to 1.5 m/s (Koch 2014).



**Fig. 10** (a) Signal of a natural hidden crack in ferritic steel detected by induction thermography at 1500 Hz excitation frequency. Image size 40 mm × 45 mm. (b) The same area imaged by induction thermography at 100 kHz. Image size 13 mm × 30 mm. (c) Metallographic cross-section along the arrow



**Fig. 11** View of the DEFECTOVISION® IR testing system. (Source: Institut Dr. Foerster GmbH & Co. KG, with permission)

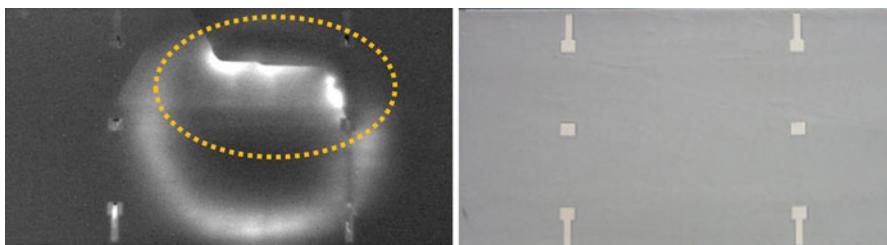
Hand-held induction systems were reported that allow mobile testing. A system developed by Siemens AG CT consists of a hand-held unit with inductor, small IR camera, and display which is connected to a rack with the power electronics (Vrana et al. 2008). Another portable solution is the ITS100 induction thermography system by Qi2/Quest Integrated, LLC. The system consists of a hand-held heating unit and a separate power generator.

## Semiconductors

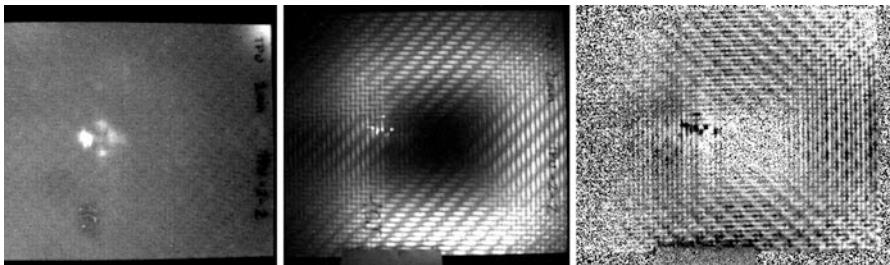
Thermographic techniques have proven to be very useful for characterization of localized defects of different nature in solar cells (Breitenstein et al. 2003). Induction thermography can also be used for testing of solar cells, as silicon is a moderate electrical conductor. Silicon is a good thermal conductor and belongs to material group IV (Table 1). One task is to detect cracks in the silicon in production processes. At usual induction frequencies, the electromagnetic skin depth is of the order of a few cm. Therefore currents are generated in the full volume of a typical cell. As silicon is quite transparent in the thermal infrared, the radiation detected is usually originating from the back-side metallization. Cracks in the brittle material can be long, but they are visually difficult to recognize. Currents that were induced by a coil on the back-side of the cell have to circumvent the cracks and produce a rapidly decaying but strong thermal signal. For the measurement shown in Fig. 12 was obtained in transmission arrangement. The coil was placed behind the sample. A 70 ms long burst pulse at 180 kHz was applied. The image shows a long crack accompanied by thermal contrasts. The thermal pattern of the circular coil is visible in the image.

## Carbon Fiber-Reinforced Polymer (CFRP)

Carbon fiber-reinforced polymers (CFRP) are electrically conducting due to their content of carbon fibers. Beside the conduction in fibers and at their contact points, displacement currents are flowing in the dielectric matrix. According to Table 1, the skin depth at typical induction frequencies is of some centimeters. Therefore,



**Fig. 12** Left: Thermographic image of a solar cell with crack (marked) Image size 85 mm × 155 mm. Right: Corresponding photo of the back side of the solar cell



**Fig. 13** Left: Detection of an impact damage in CFRP by flash excited thermography. Image size is 100 mm × 100 mm. The same area was imaged by induction thermography with lock-in excitation. Middle: amplitude image. Right: phase image (at 9 Hz modulation frequency)

components are often more or less homogeneously heated. But the current paths in the compound can be very different from that of a metal (Guo et al. 2017), in particular in uni-axial CFRP. Although the heating efficiency can be very good, flash thermography is often better to detect typical impact damage. Figure 13 shows a comparison. The impact damage is revealed clearly in the flash thermographic image. In the corresponding induction thermographic image, the contrast is dominated by the pattern of the carbon fibers in the proximity of the induction coil, which was located behind the CFRP plate. The induction frequency was about 250 kHz. The amplitude and the phase contrast obtained suggest that inductive excitation detects predominantly the fiber damage and less the inner delaminations, which is understandable when considering the largely homogeneous heating over the plate thickness.

In order to reduce the skin depth significantly, experiments were performed at frequencies of 10–30 MHz (Lehtiniemi and Hartikainen 1994) or 52 MHz (Netzelmann et al. 2016b). In the frequency range of some tens of MHz, the skin depth can be reduced to the order of a millimeter.

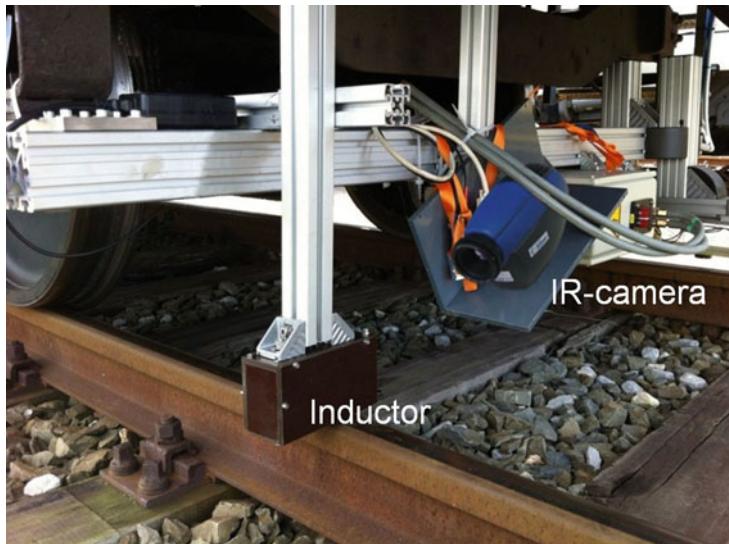
## Railway Component Testing

### Rail Surface Defects

Rail surface defects may be caused by rolling contact fatigue of the wheel and rail head.

In rails, crack-like defects (e. g. squats) may occur that are entering the rail head at a small angle. For rail testing, highly developed test cars equipped with ultrasound and eddy current devices have been in operation for many years. In order to check the possibility of induction thermography to find such surface defects in rails, a first preliminary experiment was performed on a German railway test site (Netzelmann et al. 2016a).

After tests in the laboratory and simulation work, a measurement system for induction thermography was adapted for the application in the test car (Fig. 14). An infrared camera pointed at the top surface of the rail and an inductor were mounted



**Fig. 14** Thermographic set-up for rail surface defect testing mounted under a test car

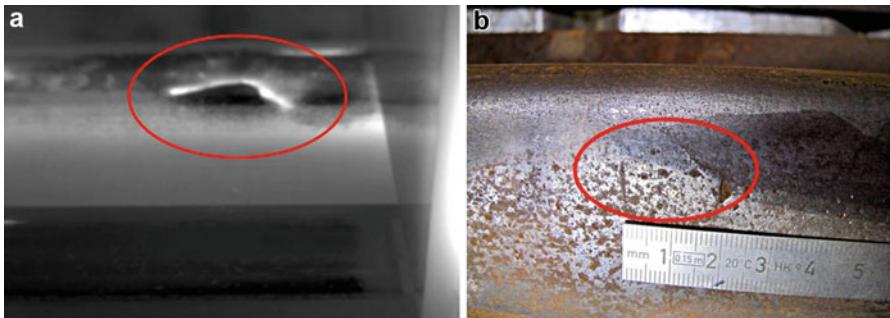
under the car close to the rail surface. The induction generator was operated in cw permanent mode and therefore required water cooling. The induction generator and the cooling equipment as well as the data recording were located in the car. The inductor is located inside a box. The camera can detect the defect contrast after it has left the area hidden by the inductor box. The time delay is acceptable due to the relatively high speed of movement.

A thermographic indication of a natural crack-like surface defect taken from the thermography image sequence without further processing is shown in Fig. 15a. It was recorded at a test car speed of 2 km/h. The defect has just left the inductor box, which appears on the right side in Fig. 15a. Measurements were performed at different speed of the test train. The crack contrast decreased with train speed, but could be detected well up to a speed of 15 km/h.

Using properly designed inductors, defects should be not only detectable on the top of the rail, but also in the curved regions between rail top and rail web. Here, other type of sensors would fail due to uncontrollable relative lateral movements between car and rail.

### Wheel Surface Defects Detected by a Fully Automated Testing System

The flexibility of camera based solutions and the proven sensitivity of induction thermography to surface defects is also the basis for a demonstrator for thermographic crack detection on railway wheels, which has been set up in Fraunhofer IZFP (Fig. 16) (Ehlen et al. 2016). Wheels are tested directly after production. The infrared camera and the inductor are moved by a robot. The wheel is fixed on a rotation stage where it can be scanned in concentric tracks over the hub, disk, and tread of the wheel. The robot keeps the required distance and direction of the arrangement of



**Fig. 15** (a) Thermographic image of a rail surface defect obtained at a speed of 2 km/h. Image size is about 70 mm × 93 mm. (b) Photo of the defect area taken in the laboratory



**Fig. 16** Demonstrator set-up for surface defect testing on railway wheels by induction thermography using a robot

inductor and camera. Thermographic image acquisition and image processing including defect recognition are performed in parallel. Results are presented in polar plots with markings of the detected defects.

Wheels with artificial test defects and with natural defects were tested. The present detection limits are cracks of 0.1 mm depth and 3 mm length. The sensitivity of the system can compete with that of magnetic particle testing. The total testing time for a wheel is below 10 min.

## Comparison of Induction Thermography and Magnetic Particle Testing

Standard NDT techniques for testing components for surface defects are magnetic particle testing (MT) and liquid penetrant testing (PT). Ferritic steel parts are often manufactured in mass production for the automotive industry. Therefore, in this industrial branch, there is big interest for fast, fully automated NDT solutions.

For MT, the testing agent is first sprayed on the component. The agent consists of a suspension of small ferromagnetic particles. By an external magnetic field or by a magnetic field generated by an external current, a magnetic flux with a strong component parallel to the surface is generated. At cracks, the magnetic flux is diverted into the region outside the component surface. The magnetic particles aggregate at the position of the maximum gradient of the magnetic field, which is the opening of the crack. Often, a fluorescent testing agent is used, where a dye is coupled to the magnetic particles. The dye emits light under excitation by a UV lamp and thereby marks the crack visible for the testing person. After the evaluation of the crack indication, the magnetic field is switched off. Then the testing agent usually has to be flushed off the surface. Disadvantages of this technique are:

- The evaluation of the test results depends on the inspector and his or her concentrated attention during the testing process.
- The differentiation between false indications and real defects is very difficult and requires an experienced inspector.
- Additional effort arises due to the necessity to remove the magnetic particle testing fluid. Sometimes demagnetization of the component is required.
- The chemicals involved are partly toxic and costly to dispose. UV radiation can be harmful to health.
- There are no digital raw data available for analysis after the surface crack detection process.

In comparison to MT, the main advantages of induction thermography are:

- The technique is working without direct contact, extensive and fast.
- Induction thermography allows detection of shallow subsurface cracks.
- By analysis of the phase signal profiles over the crack, an estimation of the crack depth is possible.
- It is not necessary to treat the surface with a couplant or particle suspension. Disposal costs and environmental hazards are avoided.
- It is usually not necessary to demagnetize a test object after testing.
- Induction thermography allows one to test the components in a fully automated objective way according to defined criteria.

## Standardization

There is considerable interest in industry on induction thermography. Up to now, however, introduction of this technique was impeded by missing standardization. A work around complex individual qualification by expertise or validation is possible. As this process is too costly and tedious in particular for small and medium enterprises, general standardization of induction thermography is desired.

In the last years, old standards on thermography were updated according to the progress of the state-of-the art. There is now a new European standard EN 16714 Nondestructive testing – Thermographic testing, containing the topics “General principles” (part 1), “Equipment” (part 2) and “Terms and definitions” (part 3). There will be a completely new European standard on active thermography based on DIN 54192. It will be followed by a European basic standard for induction thermography based on the German standard DIN 54183, which has been released at the end of 2017.

---

## Summary and Outlook

Induction thermography can be applied to a wide range of materials that exhibit at least some electrical conductivity. Its ability to detect both surface cracks and hidden cracks close to the surface is very attractive, in particular for inspection of ferritic steel. The advantage of induction thermography compared to magnetic particle testing lies in the noncontact operation avoiding particle solutions and chemicals, as well as the ability to offer fully automated testing and well-documented results.

Cracks in ferritic steel can usually be detected with good signal/noise ratio. Progress has been obtained in determining the defect depth from the thermographic data.

An application with difficulties is crack detection on highly reflecting aluminum alloys and other materials with highly reflecting surfaces and low emissivity. Simulations and experimental evidence show that the contrast obtained from a crack in aluminum may be one to two orders of magnitude smaller than that of magnetic steel.

A central requirement for an alternative testing technique, however, is at least equal detection sensitivity and reliability as the existing standard. Although present results show that relevant cracks can be detected reliably, there is a demand for research on long-term stability and reliability in mass production applications.

More pilot applications and upcoming application standards will increase the acceptance of the technique.

---

## References

- Balaji L, Balasubramanian K, Krishnamurty C (2013) Induction thermography for non-destructive evaluation of adhesive bonds. In: Review of progress in quantitative nondestructive evaluation, vol 39, AIP conference proceedings 1511, pp 579–586
- Bamberg J, Erbeck G, Zenzinger G (1999) Eddy-Therm: Ein Verfahren zur bildgebenden Prüfung metallischer Bauteile. ZfP-Zeitung 68:60–62

- Bowler N (2006) Frequency-dependence of relative permeability in steel. *Rev of Quant NDE* 25:1269–1276
- Breitenstein O, Rakotoniaina J, Al Rifai M (2003) Quantitative evaluation of shunts in solar cells by lock-in thermography. *Prog Photovolt Res Appl* 11:515–526
- Carslaw H, Jaeger J (1959) Conduction of heat in solids. Clarendon Press, Oxford, p 80
- Ehlen A, Netzelmann U, Lugin S, Finckbohner M, Valeske B, Bessert S (2016) Automated NDT of railway wheels using induction thermography. In: Proceedings of the 55th annual conference of the British Institute of non-destructive testing, Nottingham
- Guo J, Gao X, Toma E, Netzelmann U (2017) Anisotropy in carbon fiber reinforced polymer (CFRP) and its effect on induction thermography. *Nondestr Test Evaluat Int* 91:1–8
- He Y, Tian G, Pan M, Chen D (2014) Impact evaluation in carbon fiber reinforced plastic (CFRP) laminates using eddy current thermography. *Compos Struct* 109:1–7
- Heath D, Winfree W (1990) Quantitative thermal diffusivity imaging of disbonds in thermal protective coatings using inductive heating. In: Thompson DO, Chimenti DE (eds) Review of progress in quantitative nondestructive evaluation, vol 9. Plenum Press, New York, pp 577–584
- Jäckel P, Netzelmann U (2013) The influence of external magnetic fields on crack contrast in magnetic steel detected by induction thermography. *QIRT J* 10:237–247
- Koch S (2014) Non-destructive testing of bars by inductive heat-flux thermography. Millenium Steel India, pp 140–142
- Kremer K J (1984) Das THERM-O-MATIC-Verfahren – Ein neuartiges Verfahren für die Online-Prüfung von Stahlerzeugnissen auf Oberflächenfehler. In: Proceedings of the 3rd European conference in nondestructive testing, Florence, 15–18 October 1984, pp 171–186
- Lehtiniemi R, Hartikainen J (1994) An application of induction heating for fast thermal nondestructive evaluation. *Rev Sci Instrum* 65:2099–2101
- Liang T, Ren W, Tian GY, Elradi M, Gao Y (2016) Low energy impact damage detection in CFRP using eddy current pulsed thermography. *Compos Struct* 143:352–361
- Netzelmann U (2006) German Patent DE102006050025B3
- Netzelmann U, Walle G (2008) Induction thermography as a tool for reliable detection of surface defects in forged components. In: Proceedings of the 17th World conference on nondestructive testing, 25–28 Oct 2008, Shanghai, China
- Netzelmann U, Walle G, Ehlen A, Lugin S, Finckbohner M, Bessert S (2016a) NDT of railway components using induction thermography. In: AIP conference proceedings 1706, 150001
- Netzelmann U, Walle G, Lugin S, Ehlen A, Bessert S, Valeske B (2016b) Induction thermography: principle, applications and first steps towards standardization. *QIRT J* 13:170–181
- Oswald-Tranta B (2004) Thermoinductive investigations of magnetic materials for surface cracks. *QIRT J* 1:33–46
- Oswald-Tranta B (2018) Induction thermography for surface crack detection and depth determination. *Appl Sci* 8:257
- Riegert G, Zweschper T, Busse G (2004) Lockin thermography with eddy current excitation. *QIRT J* 1:21–32
- Tang B, Hou D, Hong T, Ye S (2018) Influence of the external magnetic field on crack detection in pulsed eddy current thermography. *Insight* 60:240–246
- Tsopelas N, Siakavellas N (2011) Experimental evaluation of electromagnetic-thermal non-destructive inspection by eddy current thermography in square aluminum plates. *NDT & E Int* 44:609–620
- Vrana J, Goldammer M, Baumann J, Rothenfusser M, Arnold W (2008) Mechanisms and models for crack detection with induction thermography. In: Review of progress in quantitative nondestructive evaluation, vol 27, AIP conference proceedings 975, pp 475–482
- Walle G, Netzelmann U (2006) Thermographic crack detection in ferritic steel components using inductive heating. In: Proceedings of the 9th ECNDT Berlin, 25–29 Sept 2006, DGZfP Berichtsband BB 103
- Walle G, Valeske B, Netzelmann U (2009) Eine thermische Prüftechnik zur Oberflächenrissprüfung leitfähiger Materialien. *Materialprüfung* (9):593–602

- Walle G, Netzelmann U, Stumm C, Valeske B (2012) Low frequency induction thermography for the characterization of hidden cracks in ferromagnetic steel components. In: Proceedings of the 11th international conference on quantitative infrared thermography (QIRT), 11–14 June 2012, Naples, Italy, paper 218
- Wang Y, Gao X, Netzelmann U (2018) Detection of surface cracks in metals under coatings by induction thermography. In: Proceedings of the 14th quantitative infrared thermography conference, Berlin 25–29 June 2018, DGZfP BB 167
- Wilson J, Tian G, Abidin I, Yang S, Almond D (2010) Pulsed eddy current thermography: system development and evaluation. *Insight Non-Destr Test Cond Monit* 52:87–90
- Zenzinger G, Bamberg J, Satzger W, Carl V (2007) Thermographic crack detection by eddy current excitation. *Nondestruct Test Evaluat Int* 22:101–111

---

## **Part VIII**

### **Special Techniques for Signal and Data Analysis**



# Signal Processing for NDE

42

Masoud Vejdannik, Ali Sadr, Victor Hugo C. de Albuquerque, and  
João Manuel R. S. Tavares

## Contents

Introduction .....	1526
Applications of Signal Processing to NDT .....	1527
Nondestructive Inspection Systems .....	1530
Intelligent Processing of NDT Signals .....	1530
Preprocessing .....	1532
Feature Extraction Methods .....	1534
Feature Reduction Methods .....	1538
Classification Methods .....	1539
Conclusion .....	1540
References .....	1541

### Abstract

Testing and evaluating of industrial equipment using nondestructive tests are fundamental steps in the manufacturing process. The complexity and high costs of manufacturing industrial components require examinations in some way about the quality and reliability of the specimens. However, it should be noted that in order to accurately perform the nondestructive test, in addition to

M. Vejdannik · A. Sadr

School of Electrical Engineering, Iran University of Science and Technology (IUST), Narmak, Tehran, Iran

e-mail: [masoud.vejdannik@gmail.com](mailto:masoud.vejdannik@gmail.com); [sadr@iust.ac.ir](mailto:sadr@iust.ac.ir)

V. H. C. de Albuquerque

Programa de Pós Graduação em Informática Aplicada, Universidade de Fortaleza (UNIFOR), Fortaleza, Ceará, Brazil

e-mail: [victor.albuquerque@unifor.br](mailto:victor.albuquerque@unifor.br)

J. M. R. S. Tavares (✉)

Instituto de Ciéncia e Inovação em Engenharia Mecânica e Engenharia Industrial, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

e-mail: [tavares@fe.up.pt](mailto:tavares@fe.up.pt)

theoretical knowledge, it is also essential to have experience and care, which require special courses and experience with theoretical education. Therefore, in the traditional methods, which are based on manual testing techniques and the test results depend on the operator, there is a possibility of an invalid inference from the test data. In other words, the accuracy of conclusion from the obtained data is dependent on the skill and experience of the operator. Thus, using the signal processing techniques for nondestructive evaluation (NDE), it is possible to optimize the methods of nondestructive inspection, in other words, to improve the overall system performance, in terms of reliability and system implementation costs.

In recent years, intelligent signal processing techniques have had a significant impact on the progress of nondestructive assessment. In other words, by automating the processing of nondestructive data and signals, and using artificial intelligence methods, it is possible to optimize nondestructive inspection methods, hence improving the overall system performance in terms of reliability and implementation costs of the system. This chapter reviews the issues of intelligent processing of nondestructive testing (NDT) signals.

---

## Introduction

Nondestructive testing is an integral part of modern technology and has become a critical component of engineering. The use of this technology in analyzing systems in terms of safety, reliability, and mission guarantees is vital. Today, the strategic sectors in defense, space, aeronautics, information technology, electronics, components of the nuclear power plant, railways and roads, consumer industries, material processing, conservation, restoration, and certification of cultural heritage use non-destructive tests as a worthwhile method.

Identification of material defects is one of the most important applications in nondestructive testing. Regarding the stages of manufacturing of engineering materials, occurrence of various abnormalities within them is possible. Depending on the considerations of the final operation, some of these defects and anomalies may be critical and complicated, and therefore their identification and classification are of great importance. Some of the methods used for this test include ultrasound testing, visual inspection, electromagnetic or eddy current testing, radiography, and magnetic particle testing. Identifying the defects which cause the catastrophic failures, and so are economically costly and include life-threatening hazards, is a clear advantage of using nondestructive testing. Hence, using these test methods has many advantages. The implementation of any inspection system involves costs, but often effective use of appropriate inspection methods results in significant financial savings.

Despite the many advantages that the use of nondestructive testing provides in identifying defects, the identification and classification of defects are often controversial, because the type of defects and their diagnosis are largely dependent on the user's experience and knowledge. The ability of the human eye to distinguish

meaningful patterns, after a period of training and experience, is unsurpassed. However, an overwhelming flow of data reduces the efficiency of the eye in extracting the relevant information to make the exact decision. In addition, in some applications of NDT, we seek to characterize the physical and mechanical properties of materials. For example, Vejdannik and Sadr (2015, 2016a, b, c) processed and classified the backscattered ultrasound signals which have been acquired from the specimens of nickel alloy, in order to investigate the microstructural properties of these materials. Certainly, this application is not feasible in the traditional way, or at least scanning electron microscopes should be used, which is time-consuming and costly. Therefore, traditional nondestructive testing methods based on empirical and innovative methods for identifying the pattern will lead to numerous problems in terms of cost, duration, and consequent unpredictable analysis, and thus will lead to contradiction and mismatch in the results.

Various signal processing tools, statistical methods, and soft computing are introduced to solve the existing problems. The use of these processing tools for evaluation and classification processes should address the problems mentioned in the previous section by improving the measurement accuracy and reducing the load imposed on the operator. In this way, the user can focus more effectively on the outcome of the evaluation. The main advantages of using this computational approach is to develop a decision support system for accurately identifying and classifying flaws and defects, which will provide a standardized and impartial performance.

Section “[Applications of Signal Processing to NDT](#)” briefly reviews the role of signal processing in NDT. In section “[Nondestructive Inspection Systems](#),” we will discuss about the fundamental components of NDT inspection system and issues of signal processing in typical automated NDE systems. Finally, section “[Conclusion](#)” summarizes this chapter, presents the conclusions, and describes the future directions for this topic.

---

## Applications of Signal Processing to NDT

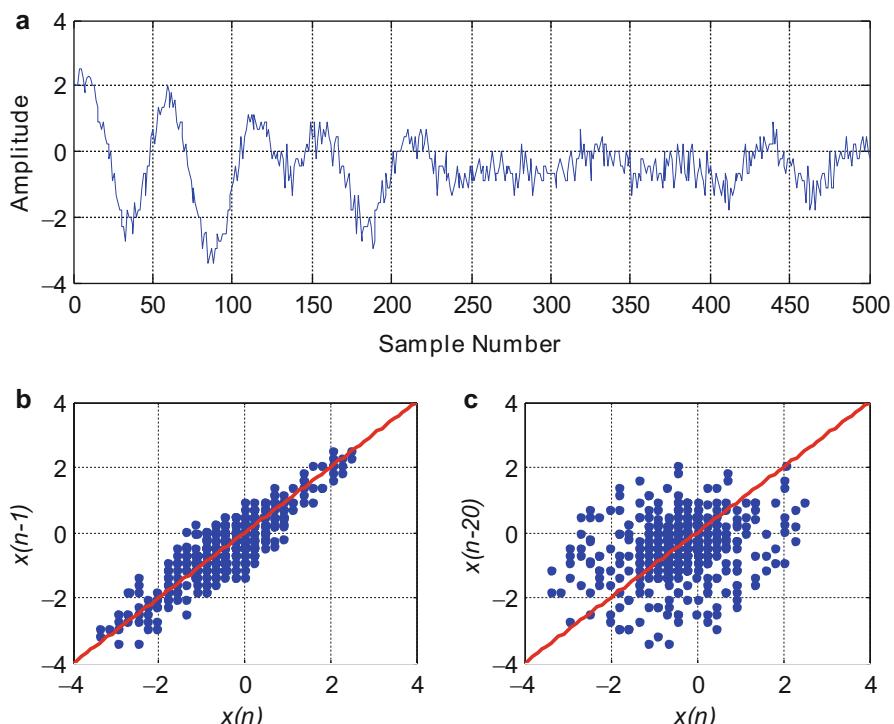
In the first years of the emergence of nondestructive assessment systems, the researchers’ efforts to improve NDT were only to increase the accuracy and resolution of inspection equipment to identify smaller defects. In other words, most of the works in the NDT domain were focused only on transducers and electrical equipment (analog signal processing). But in recent years, most of the researches and industrial applications have focused on the intelligent processing and automated processes that improve the reliability and content of information derived from the traditional methods. Many of these studies relate to the use of signal processing concepts, which have already been used successfully in other engineering areas such as radar, sonar, and geophysics.

A discrete time signal or time series is a set of observations that take place in a specified time sequence. The key characteristic of the time series is that the

observations (in other words, samples) are in order of time and the adjacent samples are correlated (Manolakis et al. 2005).

To have a better observation of the relationship between signal samples consisting of sequences of length  $l$ , we plot the points  $\{x(n), x(n+1)\}$  for  $0 \leq n \leq N-1-l$ , where  $N$  denotes the length of the signal data. This figure is known as the scattering plot. Figure 1 depicts a backscattered ultrasound signal and two associated scattering plots, which represent the correlation between adjacent samples. It can be seen that for adjacent samples, changes in the data points occur near a straight line. This implies high correlation between the adjacent samples, because each sample is followed by another sample with an almost identical amplitude. In contrast, samples that are spaced at 20-point spacing are less correlated, because data points are randomly distributed in the scattering plot (Vejdannik and Sadr 2015, 2016a, b, c).

When successive observations of the series are dependent, we may use past observations to predict future values. If the prediction is accurate, it is said that the series is a deterministic series. However, in many practical cases, we can't predict the time series exactly. Such series are called random or stochastic, and their degree of predictability is determined by the dependence between successive observations (Manolakis et al. 2005).



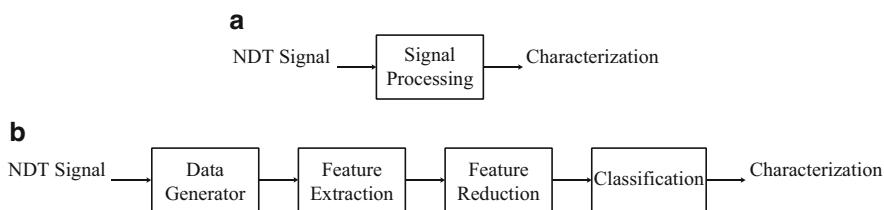
**Fig. 1** (a) A typical backscattered ultrasound signal. (b) Scattering plot for the adjacent samples. (c) Scattering plot for samples that are spaced at 20-point spacing

Signal processing is a well-known engineering term that can be defined in a broad sense as a transform that converts the signal data into useful information using a digital computer (Fig. 2a). According to the above, it is observed that most of the theoretical results in deterministic signal processing are inadequate to deal with NDT domain issues. Indeed, in many applications, the associated signal appears to be somewhat randomized, in the sense that the rerun of the test process does not result in the same observation. This is due to disturbances that the user cannot completely overcome. The disturbances can be of different types: noise or distortion effects on received signals, errors associated with measuring devices, and so on. Therefore, a state of confusion and uncertainty arises in the observations. The theory of probability and random processes provides an appropriate framework which can model this lack of knowledge. Finally, using these methods (statistical and intelligent signal processing), it is possible to take steps toward automation or, in other words, increase the accuracy of results.

In this chapter, in order to automate the NDE process in different applications, such as automatically characterize the mechanical and metallurgical properties of materials, the defects and heterogeneities within them, as well as determine the state and life duration of industrial equipment, we are going to review the different component parts of the nondestructive signal processing and classification system (Fig. 2b). This system consists of four basic blocks: data generator (preprocessing), feature extraction, feature reduction, and classification. The role of the feature extraction step is to generate a relevant set of parameters that can reveal the true nature of the context data set, which consists of a large number of data (signals). The purpose of reducing the features is to select the most relevant features for recognizing patterns. Indeed, this step will reduce the computational requirements. Also, the classification step helps in deciding which class each instance belongs to.

By analyzing the various applications of intelligent signal processing techniques in the NDT domain, it can be seen that many problems in this domain are solved by signal processing. The advantages of using signal processing for NDE include (Singh and Udpa 1986):

- Improved inspection reliability
- Improved detection of defects
- Improved characterization of defects (Nunes et al. 2013)



**Fig. 2** (a) NDT signal processing system. (b) Four basic building blocks of NDT signal processing system

- Generating useful information for quantifying the lifetime of structures (e.g., mechanical properties of materials) (Albuquerque et al. 2015)
- Generating information related to the processes. For example, processes such as thermal aging and phase transformation are monitored and controlled by NDT sensors and signal processing

---

## Nondestructive Inspection Systems

The way in which NDT signals are acquired makes a significant contribution to the overall success of the signal processing system (automated nondestructive inspection system). Failure to pair the characteristics of the signal acquisition system with the transducer can lead to irretrievable loss of information, so that subsequent attempts to recover data through signal processing methods will not be effective. For example, most ultrasound inspection equipment suffers from a low signal-to-noise ratio. This effect is caused by issues such as noise generated by acoustic and electronic sources (Chen et al. 1999). This is much more complicated in applications that use the backscattered ultrasound signals.

Therefore, critical issues in obtaining a signal, such as sampling frequency and the number of quantization levels, should be tested carefully. If signals are bandwidth-restricted, the Whittaker-Shannon sampling theory dictates that the signal must be sampled at a rate equal to at least twice the highest frequency component of the signal. Failure to follow this rule can lead to an aliasing error, which can corrupt the signal information. In typical applications, signals are sampled at a rate of two to five times the Nyquist rate to minimize the aliasing effects.

The limited number of quantizer bits leads to SNR degradation. It can be shown (Oppenheim and Schafer 2014) that for a uniform quantizer with  $b$  bits, SNR is obtained by

$$\text{SNR} = (6.02b + 10.79 + 10\log_{10}\sigma_x)\text{dB}$$

where  $\sigma_x$  is the signal variance. This relationship states that when the number of quantum bits increases, the SNR increases by about 6 dB.

Therefore, an essential requirement is to design a processing system that maximizes the capability and performance of the nondestructive inspection system. In the following, it is attempted to find out the advantages and disadvantages of various methods of intelligent signal processing in the NDT by reviewing the researches carried out in this topic.

## Intelligent Processing of NDT Signals

A signal processing system is called intelligent if it can extract the maximum information at any time from its input signal, in a nonstationary environment with unknown statistical characteristics. To enable such capability, the system must be

able to identify, exploit, and process a series of features. These features are non-linearity, adaptivity, and robustness, among others (Haikeyin 2001). To provide such requirements, inevitably, along with various signal processing tools, we must use artificial intelligence and soft computing (a set of computational methods including neural networks, fuzzy logic, and metaheuristic optimization algorithms such as genetic algorithm).

So far, we have become familiar with the advantages of using computer-based systems in analyzing nondestructive test data rather than manual analyses that always involve human error in its results. Now, it's worthwhile to compare soft and hard computing to find out the advantages and disadvantages of using these methods in analyzing NDT data and signals. In Table 1, a comparison is made between the advantages and disadvantages of soft computing and hard computing.

It can be seen from Table 1 that in many cases associated with NDT data, the use of hard computing is not feasible or there will be no satisfactory results. For example, consider the ultrasound testing case: the most important issues and problems with these signals are ambiguity, incompatibility, and high sensitivity (sensitivity to confusion in the data). For example, when faced with the delamination defects (Simas Filho et al. 2013), or defects caused by a row of holes (for example, rivet holes in thin metal sheets (Liu et al. 2013), the reflected echoes from these defects interfere together and lead to a condition of ambiguity and vagueness in making the right decision. In these conditions, fuzzy rules can create the required compatibility with ambiguity and sensitivity by generating a fuzzy model. On the other hand, in many cases, due to the special characteristics of a material and the impossibility of making an appropriate coupling between the surface of the specimen and the probe, a large amount of noise is integrated over the received signals. However, by generalizing the neural networks through learning and due to the distributed structure of these networks, they can significantly deal with the sensitivity of these signals. So far, there have been many reports on the ability of neural networks to deal with signal

**Table 1** Comparison of soft and hard computing

Hard computing	Soft computing
Input values must be accurate and definite	It can also work with vague and noisy data
The resulting answers are accurate	The resulting answers are approximate
Calculations are done sequentially	Parallel processing and computing is possible
The problem is solved algebraically	A stochastic condition is involved in solving problems
Can use only two-valued logic	Can use multiple-valued logic
To solve problems, an exact analytical model and high computing time are required	The main model is formed by the help of the human mind
To solve specific issues, reprogramming is required	Only existing apps should be upgraded
Inaccuracies are considered inappropriate	In order to gain more flexibility, they bring approximation and inaccuracy into the problem

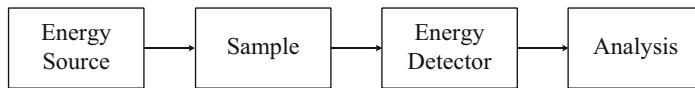
noise and lack of disturbance on the output (the decision has been taken) of these networks (Farley et al. 2012).

One of the other problems we have encountered in analyzing NDT data is the high dimensionality and inconsistency; to solve such problems genetic algorithms (evolutionary computation) can be a very convenient and efficient method (Su et al. 2015; Acciani et al. 2010). However, it should be noted that the theory of rough sets can be used to reduce the rules. Another important note is that hard computing methods cannot adapt to the complexity of nondestructive tests. For example, characterizing the type of defect among several possible defects makes such a complexity so that hard computing methods fail to analyze them (Nunes et al. 2013; Simas Filho et al. 2013; Farley et al. 2012; Su et al. 2015; Baker and Windsor 1989, 1992; Margrave et al. 1999; Santos and Perdigão 2001). However, a proper combination of methods, such as hybrid neurofuzzy networks, can provide an appropriate and efficient solution. For example, Vejdannik and Sadr (2016b, c) optimized the PNN (probabilistic neural network) and RBF (radial basis function) neural networks using the bees algorithm, and so they were able to improve the accuracy of classifying ultrasound signals for characterizing the microstructures within the materials.

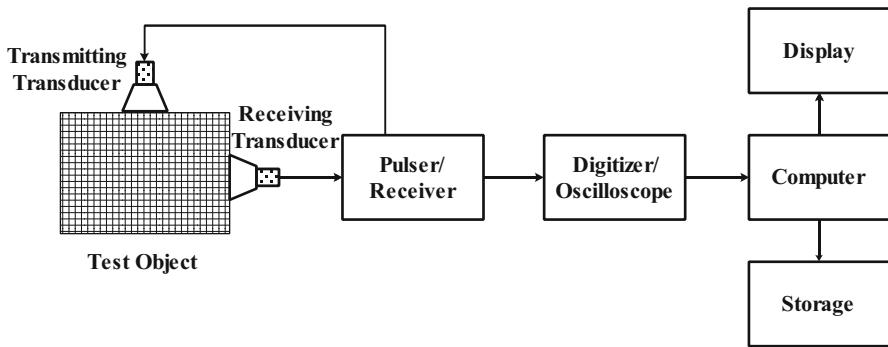
Other works of our research group evaluated the potentials of several machine learning techniques combined with various methods of feature extraction to characterize automatically the phase transformations using ultrasonic signal processing. Overall, the results achieved evidence that adopted algorithms were robust in terms of accuracy and processing time. For example, Silva et al. (2016a, b, c, d) applied an induced magnetic field on a duplex stainless steel to characterize its microstructures, Albuquerque et al. (2010a), Freitas et al. (2011), and Silva et al. (2009) analyzed the behavior of an ultrasound signal to evaluate the spinodal decomposition mechanism in duplex stainless steels under different aging thermal conditions. Albuquerque et al. (2010b, 2012) applied the concepts of NDT-based signal processing to characterize the mechanical properties of two novel materials, such as Nb-bearing nickel-based superalloy weld overlays and CuAlBe shape memory alloy, as well as in plain carbon steel (Freitas et al. 2010). In addition to one-dimensional signal processing, our group has used image analysis and processing techniques to perform nondestructive inspection in metallic and composite materials, as can be seen in Albuquerque et al. (2008, 2009, 2010a, 2011); Papa et al. (2013); Durão et al. (2012a, 2013); Papa et al. (2012); Durão et al. (2010); Albuquerque et al. (2010c); Kuncheva (2014); Cau et al. (2006).

## Preprocessing

One of the most important tasks and reasons for using signal processing algorithms in nondestructive evaluation of materials is the automatic characterization of defects within them. Before proceeding to continue the discussion, it is necessary to overview the general scheme for characterizing the flaws (Fig. 3). The energy is transmitted through a transducer into the material or the test object and interacts with



**Fig. 3** General model for NDE inspection



**Fig. 4** Basic ultrasonic test setup

the defect/material. The intensity of this interaction is a function of the type of excitation source and the properties of the material under test. On the other hand, the response signal that reflects the interaction of the energy/material is received through a receiver transducer and displayed as the output of the system. An ultrasonic test setup is a good example for demonstrating this scheme (Fig. 4).

The main challenge in this NDT scenario is to estimate the characteristics of the defects and properties of the material from the excitation and response signals. Most of the data recording and analyzing methods are very simple and basic, and thus have a low reliability. The use of signal processing can minimize the user dependencies by automatically capturing and analyzing the data. It can therefore improve the reliability of the NDT inspection system.

One of the most important goals of using signal processing in NDT is to improve the separation capability for inspecting isotropic and uniform materials, as well as materials which led to the complexity when tested by the traditional NDT techniques. This includes materials with large particles, impedance mismatching, and so on. For example, Nunes et al. (2013; Albuquerque et al. 2015) to investigate the microstructural properties of nickel alloy, which was exposed to thermal aging for 10, 100, and 200 h at 650 and 950 degrees Celsius, analyzed the backscattered ultrasound signals. Due to the microstructural changes and phase transformation in the specimens, the backscattered signals were contaminated with noise.

NDT signals are often corrupted by noise and artifacts which are generated by the transducers, tools, and other components inside the measuring equipment. This effect can be minimized by a proper signal conditioning. When the noise spectrum is accurately characterized and is outside the signal spectrum, the SNR can be improved by designing a simple filter. In practice, both of these simple methods

are inadequate in themselves, and so, more complicated methods are required. For example, Vejdannik and Sadr ([2016a, b](#)) used higher-order statistical methods in ultrasound signal processing which can implicitly remove the Gaussian noise in these signals.

## Feature Extraction Methods

A fundamental step in designing each signal classification system is the selection of a relevant set of features that can discriminate the signal in the feature space. A classification algorithm will always have somewhat identical results, but providing irrelevant features to it will not be able to reveal the true nature of the context data. Data should be reduced without losing information. Usually, in order to improve the performance and increase the computational efficiency, reducing the dimensionality of the input data is desirable, and in many cases, this is achieved by applying a series of additional and appropriate transforms. Feature selection must fulfill a fundamental criterion; features must retain all and only the significant information that is included in the data. This requirement has strong arguments about what transform should be used in the process of feature extraction. Finding the best features is a very important task and can usually be realized only through the process of trial and error.

Hence, an essential step in the nondestructive inspection systems (especially in portable and inexpensive systems) is the feature reduction step. The main reason for generating (extracting) features using the transforms is that if a proper transform is used, the redundant information will be removed from the original information. Redundant information is usually found in a set of information obtained by the measuring equipment. For example, the properties of backscattered ultrasound signals (these signals are random, nonlinear, nonstationary, and non-Gaussian) are such that the selection of features that contain the useful information from these signals is difficult and complicated. The advantage of using transforms is that the obtained information is relatively uncorrelated; thus, there is no redundant information ([Kuncheva 2014](#)).

Therefore, in choosing a feature extraction method for NDT data, it is necessary to pay attention to these points: firstly, features should be selected that maintain critical information and discard any unnecessary data. Secondly, the method used should have a high energy packing capability in order to reduce the computational requirements and subsequently reduce time and costs. In Table [2](#), a comparison is made between the common feature extraction techniques used in the analysis of nondestructive tests.

So far, there have been many reports of the success of all these methods, but each of these methods has advantages and disadvantages that a more detailed examination of them can be useful and inspiring in achieving a more efficient and more robust system. One of the most important and essential issues which must be considered in the selection of feature extraction method, especially in nondestructive tests, is the computational complexity of the method used, because most NDT inspection

**Table 2** Advantages and disadvantages of conventional feature extraction methods in NDE

Feature extraction method	Advantages	Disadvantages
Karhunen-Loeve transform	Features are completely uncorrelated	Depends on the application High computational complexity
Singular value decomposition	It has a very good packing efficiency Basis vectors are the result of an optimization process	It works on a fixed matrix. So, it is not suitable for working with adaptive algorithms Depends on the application The relatively high computational complexity is due to the optimization of $O(N^3)$
Discrete Fourier transform	Moderate computational complexity $O(N^2)$ or relatively low for fast Fourier transform $O(\log_2 N)$	It has a low data packing efficiency
Discrete cosine transform	A fast algorithm with relatively low complexity $O(\log_2 N)$ Very good packing efficiency	It requires an extra quantization step in the output
Haar transform	Multiresolution analysis tool Moderate energy packing efficiency	Frequency response of Haar filters is not ideal
Discrete wavelet transform	Multiresolution analysis tool Relatively high energy packing efficiency A fast algorithm with relatively low complexity $O(N)$	Shift-variant Aliasing Oscillation

equipment are designed and manufactured to provide more mobility for the operator and can be used in various industrial situations (Simas Filho et al. 2013). Also, if the desired transform has a high energy packing efficiency, that is, the useful information of the signal is limited to a small number of coefficients, the implementation of the system requires less computational requirements, which can also be effective in speeding up and reducing costs.

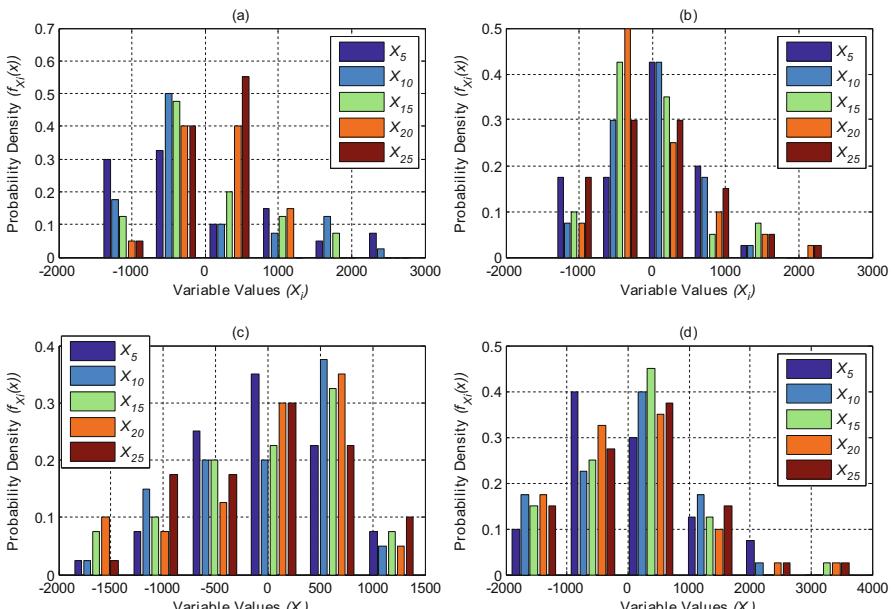
Among these methods, the SVD transform has a strong capability in information packing, and thus, NDT data can be replaced by a smaller number of its singular values. However, vectors are not constant in SVD and K-L expansions. Indeed, they are application-dependent and resulted from an optimization process. But this will increase the computational complexity. One can use the discrete Fourier transform that has less computational complexity, but instead it has a low energy packing capability. Also, less computational complexity can be provided by using the fast Fourier transform algorithm (Acciani et al. 2010, Cau et al. 2006). DCT is a member of a family of transforms that uses a fast algorithm with relatively low complexity. Because of the fact that DCT transforms the coefficients into a small domain, it provides an applicatory information packing efficiency for the NDT signals.

Among these methods, energy packing efficiency of Haar transform is rarely good but instead provides multiresolution tools. However, the wavelet transform can

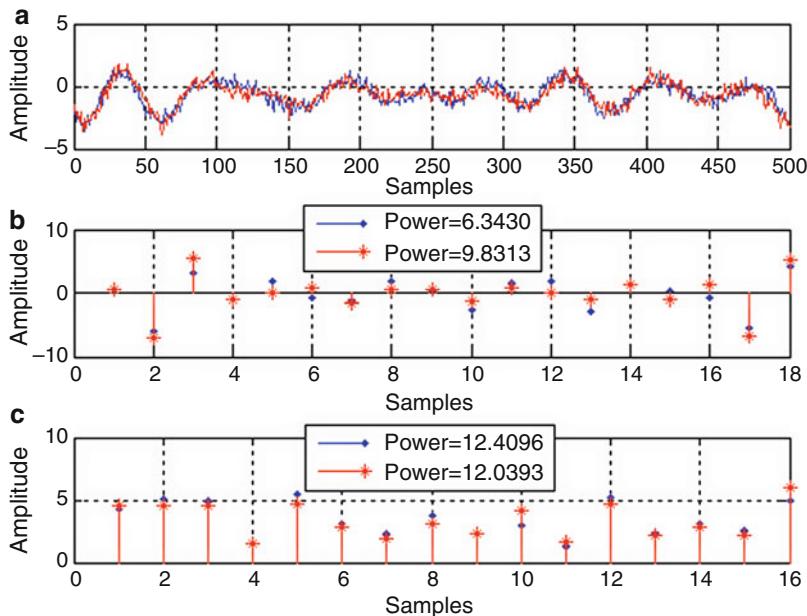
significantly improve the energy packing efficiency by choosing a proper mother wavelet. The wavelet transform has a very good localization property. In fact, the wavelet transform is used to find the moment at which an abrupt change occurs in the frequency domain at a given time.

Among the various methods used to extract features in the NDT signal processing, most successful reports relate to the use of wavelet transform (Liu et al. 2013; Farley et al. 2012; Acciani et al. 2010; Barry et al. 2015; Crouch et al. 2015; Kesharaju and Nagarajah 2015; Iyer et al. 2012; Kesharaju et al. 2014). In fact, due to the properties of most NDT signals (these signals are nonstationary) which often led to the distribution of information at specific resolution levels, energy is often concentrated in a small number of samples. Although the DWT processing method is an efficient and effective method for analyzing nonstationary signals such as ultrasound signals (Fig. 5 shows the approximate probability distribution for the five variables of the random process  $x(t)$ , with the same intervals. The unequal distribution of probabilities in these random variables indicates the variability of the probability distribution function in time, or in other words, the nonstationarity of the signal  $x(t)$ ), this method lacks the shift invariance property. Therefore, when the input signal is shifted, the energy of the wavelet coefficients varies considerably. The DTCWT processing method solves this problem by considering a complex structure (inspired by the Fourier transform).

As can be seen in Fig. 6, variations in the average power of complex wavelet coefficients are very small compared to the average power of discrete wavelet



**Fig. 5** Approximate probability distribution for the five variables of the random process  $x(t)$  (an ultrasound signal in four different microstructural classes (Vejdannik and Sadr 2016c), with the same intervals – the unequal distribution of probabilities reveals that the signal is nonstationary)



**Fig. 6** (a) Two ultrasound signals in two different classes (Vejdannik and Sadr 2016c). (b) Fifth scale detail coefficients of the (a) signals using DWT. (c) Fifth scale detail coefficients of the (a) signals using DTCWT

coefficients. Therefore, the application of complex wavelet transform is more efficient in extracting invariant features for the classification of NDT signals than discrete wavelet transform (Vejdannik and Sadr 2016c).

On the other hand, NDT signals are inherently nonlinear, so conventional linear methods in the time and frequency domain cannot clearly characterize the complexity of such signals. For this purpose, Vejdannik and Sadr (Vejdannik and Sadr 2016a) used higher-order spectral analysis methods to detect the nonlinear relationship type and phase information between frequency components to characterize the different phases in the process of phase transformation of Inconel 625 alloy. The motivation to use the higher-order spectral analysis methods are:

1. This method can eliminate Gaussian noise in the unknown spectral specifications. If a non-Gaussian signal is combined with a Gaussian noise, this noise can be eliminated using an HOS transform (this will increase the signal-to-noise ratio).
2. HOS methods maintain the phase information. For example in practice, sometimes, some of the harmonic frequencies interact with each other and cause the summation/subtraction of their frequencies to appear as a new frequency component in the power spectrum.
3. HOS methods can play a key role in identifying and characterizing nonlinearity in a system.

In addition to the above, scaling of the extracted features is also necessary, especially when the data covers different ranges. After scaling, the normalization of the data set is implemented, so all the inputs and target outputs have the same mean of zero and standard deviation of 1. The goal of normalization is to ensure that all the features have an almost equal weight and that large quantities will not prevail over other values.

## Feature Reduction Methods

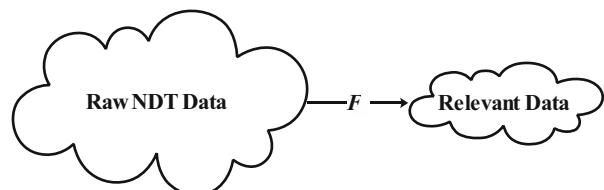
One of the main issues in classification/pattern recognition (compatible with different NDT tests) is to select an appropriate range of features, which provides independent or discriminant information required by the classifier for the correct classification of defects. Often, a large collection of potentially useful features is collected (feature extraction), and then the most appropriate and relevant features are selected by reducing the features (Fig. 7). Indeed, all extracted features are not relevant to the pattern recognition task. These features entail some kind of doubt and ambiguity on the problem, and consequently reduce the ability of identification (Duda et al. 2001).

The methods used in the analysis of nondestructive tests for feature reduction include genetic optimization algorithms (Su et al. 2015; Acciani et al. 2010; Bustillo 2014a, b), principal component analysis (Simas Filho et al. 2013; Cau et al. 2006), RS, and DFA (Nunes et al. 2013). Among these works, the RS and DFA algorithms have yielded poor results, but GA and PCA have brought successful results. Vejdannik and Sadr in (Vejdannik and Sadr 2016a) and (Vejdannik and Sadr 2015) used LDA to reduce the number of features in ultrasound signal processing and consequently ANOVA to select relevant coefficients.

In (Vejdannik and Sadr 2016d), three statistical methods PCA, LDA, and ICA were used to reduce the number of NDT features, and the results were compared with each other. Among the feature reduction methods applied in Vejdannik and Sadr 2016d), ICA performed better than LDA, and also LDA performed better than PCA. Indeed, PCA seeks to project features in the direction of the highest variability which are sorted by the eigenvectors. However, these directions do not necessarily provide the highest possible discrimination.

Like the PCA, LDA is looking for a linear combination of variables to reveal data in a better manner. However, PCA does not distinguish between different classes, while LDA explicitly tries to model the difference between the classes of data.

**Fig. 7** Mapping of raw NDT data into feature space



Indeed, LDA provides the highest discrimination between different classes and can therefore lead to a higher classification accuracy than the PCA method.

Besides, PCA and LDA methods explore the second-order statistics and thus can only eliminate the statistical correlation between features. However, the ICA method uses higher-order statistics and therefore provides statistical independence. Indeed, higher-order statistics can reveal more significant information from the features. Hence, they can provide more discriminant features and consequently higher classification accuracy. The results obtained in Vejdannik and Sadr (2016d) prove this.

## Classification Methods

Classification is referred to a procedure in which we associate an unknown pattern (feature vector) to a class. This class can be one of the classes previously specified, or created by one or more new classes with a collection of unknown patterns.

The term *supervised classification* is referred to a procedure in which a classifier design is performed using a set of training samples. In other words, the label of each class is predetermined. Then, after designing the classifier, a set of test data is fed to the classifier, and the result of classifying each sample is determined.

In pattern recognition, the training samples and the corresponding classes are not always available. In this type of problems, there is a set of  $x$  feature vectors, and the purpose of the problem is identifying the similarities and grouping or categorizing the similar vectors. This procedure is called *unsupervised classification* or clustering. A reasonable clustering algorithm produces clusters, where in each cluster, samples are close together, while the similarity between clusters is as small as possible.

The final step in designing a decision support system (automated nondestructive inspection system) is to select a classifier, so that it can, while maintaining high accuracy and speed (in comparison with the existing methods and taking into account the requirements of each particular NDT application) and low computational requirements (with regard to the type of system, including portable and fixed), provide high reliability. Because the failure to identify some defects in industrial applications is catastrophic, it leads to irreparable accidents.

The important point is that if we consider only the factor of accuracy and reliability, we cannot definitely announce the superiority of a classifier than others. But depending on the signal and the specific features that it has, each of these classifiers can achieve better result. For example, each of the MLP, RBF (Song and Schmerr 1992), PNN, FCM (Akram et al. 2014), SVM, OPF (Nunes et al. 2013), LDA (Kesharaju et al. 2014), and Bayesian (Nunes et al. 2013; Su et al. 2013) classifiers can, in turn, have successful results by taking into account the characteristics of their classes. However, the reports obtained in Nunes et al. (2013) demonstrate that the Bayesian classifier has a higher accuracy than OPF, and the OPF classifier provides a higher accuracy than the SVM classifier. It should also be noted that several papers and experiments have reported the relative superiority of MLP classifiers to SOM, LVQ, SVM, and fuzzy methods in the classification of defects.

Now, if we consider other factors such as computational requirements and speed, their fair comparison is feasible and provides useful results. Neural networks, due to their parallel structure, have a relatively high processing speed. However, they demand high computational requirements correspondingly (Karray and De Silva 2004). However, OPF requires much less computational requirements while maintaining high processing speeds (Nunes et al. 2013). SVM is also far slower than this and demands more computational requirement than OPF. On the other hand, the proposed neural networks in Vejdannik and Sadr (2016b, c) provide much higher accuracy and speed. In these neural networks, four different approaches are presented for estimating the probability density function. Also, an optimization algorithm based on the Bees Algorithm is provided to set the parameters for these classifiers. The results obtained in these processing systems indicate that the intelligent NDT signal processing can greatly improve the reliability of nondestructive inspection systems by automating the process of characterizing defects and material properties.

---

## Conclusion

Due to the rapid growth in the use of nondestructive testing to identify defects, the study of material properties, the growth of available data volumes, and the rapid increase in the methods of collecting these data, traditional and manual analysis of data is inefficient and computer-aided methods are accelerating. Novel signal processing techniques and artificial intelligence algorithms are being tested to solve various problems related to the nondestructive testing, especially ultrasound testing, so that they can provide a decision support system. This fast and user-friendly system can help industrial users and technicians to make decisions by reducing the effort, time, and, consequently the cost of classifying ultrasound test signals.

Indeed, in the traditional function, because the test results are dependent on the human user, the possibility of an invalid conclusion is very high. In other words, the accuracy of conclusion from the obtained data is dependent on the skill and experience of the examiner. Thus, by the automated processing of NDT signals and using artificial intelligence techniques, it is possible to step up the optimization of nondestructive inspection methods, namely, improving the overall system performance, in terms of reliability and system implementation costs. In this regard, due to the random, nonlinear, and nonstationary properties of the NDT signals, AI methods and statistical signal processing techniques have been able to play an effective role in solving various NDT problems. However, in the NDE domain, less attention has been paid to the statistical processing of NDT signals than other engineering areas. Therefore, according to the properties of these signals, it is desirable to focus more on this branch in the future.

**Acknowledgments** VHCA acknowledges the sponsorship from the Brazilian National Council for Research and Development (CNPq) via Grant No. 301928/2014-2. JMRST gratefully acknowledges the funding of Project NORTE-01-0145-FEDER-000022 – SciTech – Science and

Technology for Competitive and Sustainable Industries, co-financed by “Programa Operacional Regional do Norte” (NORTE2020), through “Fundo Europeu de Desenvolvimento Regional” (FEDER).

---

## References

- Acciani G, Brunetti G, Fornarelli G, Giaquinto A (2010) Angular and axial evaluation of superficial defects on non-accessible pipes by wavelet transform and neural network-based classification. *Ultrasonics* 50:13–25
- Akram N, Isa D, Rajkumar R, Lee L (2014) Active incremental support vector machine for oil and gas pipeline defects prediction system using long range ultrasonic transducers. *Ultrasonics* 54:1534–1544
- Baker A, Windsor C (1989) The classification of defects from ultrasonic data using neural networks: The Hopfield method, *NDT International*, 22(2):97–105
- Baker A, Windsor C (1992) The classification of defects from ultrasonic data using neural networks: the Hopfield method, *NDT & E International*, 25(1):46
- Barry T, Kesharaju M, Nagarajah C, Palanisamy S (2015) Defect characterisation in laminar composite structures using ultrasonic techniques and artificial neural networks. *J Compos Mater* 50:861–871
- Bustillo J, Fortineau J, Gautier G, Lethiecq M (2014a) Ultrasonic characterization of porous silicon using a genetic algorithm to solve the inverse problem. *NDT E Int* 62:93–98
- Bustillo J, Fortineau J, Gautier G, Lethiecq M (2014b) Ultrasonic characterization of electrochemically etched porous silicon. *Jpn J Appl Phys* 53:060308
- Cau F, Fanni A, Montisci A, Testoni P, Usai M (2006) A signal-processing tool for non-destructive testing of inaccessible pipes. *Eng Appl Artif Intell* 19:753–760
- Chen J, Shi Y, Shi S (1999) Noise analysis of digital ultrasonic nondestructive evaluation system. *Int J Press Vessel Pip* 76:619–630
- Crouch I, Kesharaju M, Nagarajah R (2015) Characterisation, significance and detection of manufacturing defects in Reaction Sintered Silicon Carbide armour materials. *Ceram Int* 41:11581–11591
- de Albuquerque V, Cortez P, de Alexandria A, Tavares J (2008) A new solution for automatic microstructures analysis from images based on a backpropagation artificial neural network. *Nondestruct Test Eval* 23:273–283
- de Albuquerque V, de Alexandria A, Cortez P, Tavares J (2009a) Evaluation of multilayer perceptron and self-organizing map neural network topologies applied on microstructure segmentation from metallographic images. *NDT E Int* 42:644–651
- de Albuquerque V, Tavares J, Durão L (2009b) Evaluation of delamination damage on composite plates using an artificial neural network for the radiographic image analysis. *J Compos Mater* 44:1139–1159
- de Albuquerque V, Filho P, Cavalcante T, Tavares J (2010a) New computational solution to quantify synthetic material porosity from optical microscopic images. *J Microsc* 240:50–59
- de Albuquerque V, de Macedo Silva E, Pereira Leite J, de Moura E, de Araújo Freitas V, Tavares J (2010b) Spinodal decomposition mechanism study on the duplex stainless steel UNS S31803 using ultrasonic speed measurements. *Mater Des* 31:2147–2150
- de Albuquerque V, Melo T, de Oliveira D, Gomes R, Tavares J (2010c) Evaluation of grain refiners influence on the mechanical properties in a CuAlBe shape memory alloy by ultrasonic and mechanical tensile testing. *Mater Des* 31:3275–3281
- de Albuquerque V, Silva C, Menezes T, Farias J, Tavares J (2010d) Automatic evaluation of nickel alloy secondary phases from SEM images. *Microsc Res Tech* 74:36–46
- de Albuquerque V, Silva C, Normando P, Moura E, Tavares J (2012) Thermal aging effects on the microstructure of Nb-bearing nickel based superalloy weld overlays using ultrasound techniques. *Mater Des* 36:337–347

- de Albuquerque V, Barbosa C, Silva C, Moura E, Filho P, Papa J, Tavares J (2015) Ultrasonic sensor signals and optimum path forest classifier for the microstructural characterization of thermally-aged inconel 625 alloy. *Sensors* 15:12474–12497
- de Araújo Freitas V, Normando P, de Albuquerque V, de Macedo Silva E, Silva A, Tavares J (2011) Nondestructive characterization and evaluation of embrittlement kinetics and elastic constants of duplex stainless steel SAF 2205 for different aging times at 425 °C and 475 °C. *J Nondestruct Eval* 30:130–136
- de Macedo Silva E, de Albuquerque V, Leite J, Varela A, Moura E, Tavares J (2009) Phase transformations evaluation on a UNS S31803 duplex stainless steel based on nondestructive testing. *Mater Sci Eng A* 516:126–130
- de Macedo Silva E, Leite J, de França Neto F, Leite J, Fialho W, de Albuquerque V, Tavares J (2014) Evaluation of the magnetic permeability for the microstructural characterization of a duplex stainless steel. *J Test Eval* 44:20130313
- de Macedo Silva E, Leite J, Leite J, Fialho W, de Albuquerque V, Tavares J (2016) Induced magnetic field used to detect the sigma phase of a 2205 duplex stainless steel. *J Nondestruct Eval* 35. <https://doi.org/10.1007/s10921-016-0339-7>
- Duda R, Hart P, Stork D (2001) Pattern classification. New York, NY: John Wiley & Sons
- Durão L, Tavares J, de Albuquerque V, Marques A, Magalhães A, Vieira A (2010) Tool effects on hybrid laminates drilling. *Mater Manuf Process* 25:476–481
- Durão L, Gonçalves D, Tavares J, de Albuquerque V, Marques A (2011) Comparative analysis of drills for composite laminates. *J Compos Mater* 46:1649–1659
- Durão L, Tavares J, de Albuquerque V, Gonçalves D (2013) Damage evaluation of drilled carbon/epoxy laminates based on area assessment methods. *Compos Struct* 96:576–583
- Farley S, Durodola J, Fellows N, Hernández-Gómez L (2012) High resolution non-destructive evaluation of defects using artificial neural networks and wavelets. *NDT E Int* 52:69–75
- Freitas V, Albuquerque V, Silva E, Silva A, Tavares J (2010) Nondestructive characterization of microstructures and determination of elastic properties in plain carbon steel using ultrasonic measurements. *Mater Sci Eng A* 527:4431–4437
- Haikeyin (2001) A comprehensive foundation. Prentice Hall PTR
- Iyer S, Sinha S, Tittmann B, Pedrick M (2012) Ultrasonic signal processing methods for detection of defects in concrete pipes. *Autom Constr* 22:135–148
- Karray F, De Silva C (2004) Soft computing and tools of intelligent systems design. Addison-Wesley, Harlow
- Kesharaju M, Nagarajah R (2015) Feature selection for neural network based defect classification of ceramic components using high frequency ultrasound. *Ultrasonics* 62:271–277
- Kesharaju M, Nagarajah R, Zhang T, Crouch I (2014) Ultrasonic sensor based defect detection and characterisation of ceramics. *Ultrasonics* 54:312–317
- Kuncheva L (2014) Combining pattern classifiers. Wiley, Hoboken
- Liu S, Du C, Mou J, Martua L, Zhang J, Lewis F (2013) Diagnosis of structural cracks using wavelet transform and neural networks. *NDT E Int* 54:9–18
- Manolakis D, Ingle V, Kogon S (2005) Statistical and adaptive signal processing. Artech House, Boston
- Margrave F, Rigas K, Bradley D, Barrowcliffe P (1999) The use of neural networks in ultrasonic flaw detection. *Measurement* 25:143–154
- Nunes T, de Albuquerque V, Papa J, Silva C, Normando P, Moura E, Tavares J (2013) Automatic microstructural characterization and classification using artificial intelligence techniques on ultrasound signals. *Expert Syst Appl* 40:3096–3105
- Oppenheim A, Schafer R (2014) Discrete-time signal processing. Pearson, Harlow
- Papa J, Falcão A, de Albuquerque V, Tavares J (2012) Efficient supervised optimum-path forest classification for large datasets. *Pattern Recogn* 45:512–520
- Papa J, Nakamura R, de Albuquerque V, Falcão A, Tavares J (2013) Computer techniques towards the automatic characterization of graphite particles in metallographic images of industrial materials. *Expert Syst Appl* 40:590–597

- Santos J, Perdigão F (2001) Automatic defects classification – a contribution. *NDT E Int* 34:313–318
- Silva E, Marinho L, Filho P, Leite J, Leite J, Fialho W, de Albuquerque V, Tavares J (2016a) Classification of induced magnetic field signals for the microstructural characterization of sigma phase in duplex stainless steels. *Metals* 6:164
- Silva E, Paula A, Leite J, Leite J, Andrade L, de Albuquerque V, Tavares J (2016b) Detection of the magnetic easy direction in steels using induced magnetic fields. *Metals* 6:317
- Simas Filho E, Souza Y, Lopes J, Farias C, Albuquerque M (2013) Decision support system for ultrasound inspection of fiber metal laminates using statistical signal processing and neural networks. *Ultrasonics* 53:1104–1111
- Singh G, Udupa S (1986) The role of digital signal processing in NDT. *NDT Int* 19:125–132
- Song S, Schmerr L (1992) Ultrasonic flaw classification in weldments using probabilistic neural networks. *J Nondestruct Eval* 11:69–77
- Su L, Zha Z, Lu X, Shi T, Liao G (2013) Using BP network for ultrasonic inspection of flip chip solder joints. *Mech Syst Signal Process* 34:183–190
- Su L, Shi T, Du L, Lu X, Liao G (2015) Genetic algorithms for defect detection of flip chips. *Microelectron Reliab* 55:213–220
- Vejdannik M, Sadr A (2015) Application of linear discriminant analysis to ultrasound signals for automatic microstructural characterization and classification. *J Signal Process Sys* 83(3):411–421
- Vejdannik M, Sadr A (2016a) Automatic microstructural characterization and classification using higher-order spectra on ultrasound signals. *J Nondestruct Eval* 35:1
- Vejdannik M, Sadr A (2016b) Automatic microstructural characterization and classification using probabilistic neural network on ultrasound signals. *J Intell Manuf*. <https://doi.org/10.1007/s10845-016-1225-y>
- Vejdannik M, Sadr A (2016c) Automatic microstructural characterization and classification using dual tree complex wavelet-based features and Bees Algorithm. *Neural Comput Appl* 28(7):1877–1889
- Vejdannik M, Sadr A (2016d) Application of statistical signal processing techniques to ultrasound signals for automatic microstructural characterization and classification, In: *3rd International Iranian NDT Conference*, Tehran: IRAN NDT



# Digital Image Correlation Techniques for NDE and SHM

43

Christopher Niezrecki, Javad Baqersad, and Alessandro Sabato

## Contents

Introduction .....	1546
Basics of Digital Image Correlation .....	1548
Nondestructive Evaluation and Structural Health Monitoring Using Digital Image Correlation .....	1550
NDE and SHM Using DIC in Aerospace Applications .....	1550
NDE and SHM Using DIC in Civil Applications .....	1552
NDE and SHM Using DIC in Dynamic Applications .....	1562
NDE and SHM Using DIC in Energy Applications .....	1566
NDE and SHM Using DIC in Material Testing Applications .....	1572
Other Applications and Vision-Based Systems .....	1578
Future Directions .....	1579
Conclusions .....	1580
References .....	1580

## Abstract

Monitoring and analyzing the integrity of structures, infrastructure, and machines is essential for economic, operational, and safety reasons. The assessment of structural integrity and dynamic conditions of those systems is important to ensure safe operation and achieve or even extend the design service life. Recent advancements in camera technology, optical sensors, and image processing algorithms have made optically based and noncontact measurement techniques such as photogrammetry and digital image correlation (DIC) appealing methods

---

C. Niezrecki (✉) · A. Sabato

Department of Mechanical Engineering, University of Massachusetts Lowell, Lowell, MA, USA  
e-mail: [Christopher\\_Niezrecki@uml.edu](mailto:Christopher_Niezrecki@uml.edu); [Alessandro\\_Sabato@uml.edu](mailto:Alessandro_Sabato@uml.edu); [sabatoale@gmail.com](mailto:sabatoale@gmail.com)

J. Baqersad

Department of Mechanical Engineering, Kettering University, Flint, MI, USA  
e-mail: [jbaqersad@kettering.edu](mailto:jbaqersad@kettering.edu)

for nondestructive evaluation (NDE) and structural health monitoring (SHM). Conventional sensors (e.g., accelerometers, strain gages, string potentiometers, LVDTs) provide results only at a discrete number of points. Moreover, these sensors need wiring, can be time-consuming to install, may require additional instrumentations (e.g., power amplifiers, data acquisition), and are difficult to implement on large-sized structures without interfering with their functionality or may require instrumentation having a large number of data channels. On the contrary, optical techniques can provide accurate quantitative information about full-field displacement, strain, geometry, and the dynamics of a structure without contact or interfering with the structure's functionality. This chapter presents a summary review of the efforts made in both academia and industry to leverage the use of DIC systems for NDE and SHM applications in the fields of civil, aerospace, and energy engineering systems. The chapter also summarizes the feasibility of the approaches and presents possible future directions of the measurement approach.

---

## Introduction

Aerospace, civil, and mechanical engineering structures such as bridges, buildings, pipelines, aircraft, ships, and wind turbines are routinely used despite the reality that many are aging, deteriorating, and exceeding their intended operational design life (e.g., ASCE 2017). Assessing and quantifying the condition of aging structures is essential to verify structural integrity, ensure long-term reliability, and determine when component repair or replacement should be made. A goal in industry is to move away from schedule-based maintenance to condition-based monitoring to perform assessment of factors that can jeopardize the system's performance. Identifying a strategy to detect damage for engineering systems, structures, and infrastructure is called structural health monitoring (SHM). SHM involves the observation of the targeted system over time to extract damage-sensitive features, determine the current health state, and predict future condition (Farrar and Worden 2007) which is important for damage prognosis and future life prediction. SHM plays a key role in the prevention of catastrophic failures, improving the safety of structures and infrastructure and reducing the downtime and costs associated with maintenance. Thus, it represents a method for tracking the health of an engineering system by combining damage detection algorithms with structural monitoring devices (e.g., sensors). SHM is often carried in conjunction with another closely related discipline: nondestructive evaluation (NDE) often referred to as nondestructive testing (NDT). Those techniques consist of some evaluation methods to assess the condition of a targeted system without affecting the system's functionality (Shull 2016). Both SHM and NDE techniques allow for early detection and assessment of structural damage to ensure that structures continue to meet life-safety requirements. Because human visual inspection techniques are based on the inspector's opinion, subject to variability, and labor-intensive, this makes automated computer-based monitoring systems very desirable.

Contact-based sensors are commonly used for monitoring a variety of structural systems. One of the most common practices is to record and analyze the data from a network of sensors, either passive or active, embedded or attached onto the monitored structure (Giurgiutiu and Cuc 2005). These methods include both dynamic and static analyses and have significantly improved over time. Contact-based sensors such as strain gages, accelerometers, linear voltage displacement transducers (LVDTs), inclinometers, and extensometers are commonly used for SHM and NDE applications (Doebling et al. 1996; Sohn et al. 2002). Fiber-optic sensors have proven to be valid alternatives to the conventional sensors due to their flexibility, electromagnetic interference (EMI) immunity, and scalability (Glisic and Inaudi 2008; Ye et al. 2014). However, these sensors can be difficult to implement, need wiring, are costly, require power, and once attached are generally not portable for interrogation on multiple engineering systems. For resolving these problems, several researchers proposed using wireless sensor networks (Lynch 2007; Sabato et al. 2017a). Nevertheless, even these sensors are typically not durable enough to be attached or embedded in the structure and perform measurements throughout a structure's life span which may be years or decades after its construction (when failures are more likely to occur). Furthermore, most of these sensors can only provide information at a few discrete points (Calebi 2000).

Recent technological developments have provided new NDE techniques for the assessment of different typologies of engineering systems. Radiography (Udod et al. 2016), radioactive computerized tomography (Hanke et al. 2008), radar (Pieraccini 2013), ultrasonic arrays and acoustic imaging systems (Drinkwater and Wilcox 2006), acoustic emission (Gholizadeh et al. 2015), and infrared thermography (Usamentiaga et al. 2014) have all been implemented for NDE and SHM, and each possesses their advantages and challenges. The readers are referred to the studies of Gholizadeh (2016; composite materials); Rehman et al. (2016) and Seo et al. (2015) (bridges); Drewry and Georgiou (2007), Tchakoua et al. (2014), and Zhou et al. (2014) (wind turbines); and Rizzo (2014; railroads) for further information, depending on the specific SHM and NDE applications.

New advances in camera technology, optical sensors, and image processing algorithms allowed the development of a new generation of noncontact measuring methods. Optical-based techniques such as three-dimensional digital image correlation (3D-DIC) and three-dimensional point tracking (3DPT) have become valuable tools for performing noncontact measurements and extracting structural deformations, full-field displacement and strain, geometry profiles, and modal parameters in aerospace, civil, and mechanical engineering systems. In this chapter, a review of the efforts made both in academia and industry to leverage the use of DIC systems for performing SHM and NDE analyses is summarized.

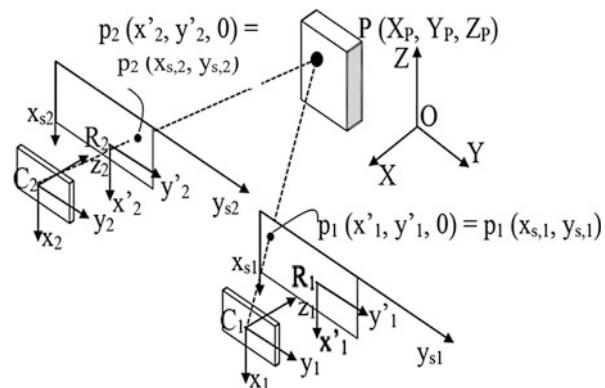
This chapter is organized as follows: Section “[Basics of Digital Image Correlation](#)” presents a description of the theoretical foundation of optically based and computer-vision-based technologies. In section “[Nondestructive Evaluation and Structural Health Monitoring Using Digital Image Correlation](#),” a detailed report of the state-of-the-art of computer-vision-based systems employed for SHM and NDE applications is discussed. Finally, future work is briefly outlined in section “[Future Directions](#),” and conclusions are drawn in section “[Conclusions](#).“

## Basics of Digital Image Correlation

Photogrammetry and vision-based techniques use photographs recorded with imaging sensors to identify coordinates of points, features, and patterns of an object and use this data to track their motion thorough different times or stages. This technique was further developed and used for deformation measurement in the early 1980s by Peters, Ranson, Sutton et al. (Peters and Ranson 1982; Peters et al. 1983; Sutton et al. 1983). Through the years, the vision-based systems have proven to be a reliable technique to measure displacements occurring in large-sized structures, evaluating mechanical stress-strain fields and extracting modal parameters.

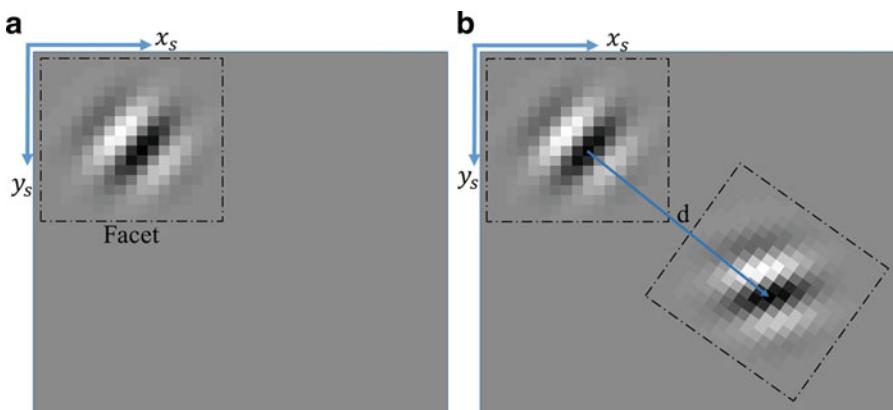
Three-dimensional photogrammetry is a noncontact, full-field, optical measuring method capable of extracting surface displacement and geometry profiles from images acquired through a synchronized stereo camera (or multi-camera) system. A generic point  $P$  on the surface of the target object can be represented by its coordinates ( $X_P$ ,  $Y_P$ ,  $Z_P$ ) in the global coordinate system  $X-Y-Z$  with an origin located at point O. A schematic of the camera setup is shown in Fig. 1. As can be seen, each camera has its coordinate system ( $x_i$ ,  $y_i$ ,  $z_i$ ) with origin at  $C_i$ , where  $i = 1, 2$ , while another plane can be defined in which the projection of the point P has coordinates ( $x'_i$ ,  $y'_i$ ,  $f'_i$ ). This plane belongs to the image or retinal system  $R_i$ , which is obtained as  $z_i$  is shifted by a distance equal to  $f'_i$ , so the point has coordinates ( $x'_i$ ,  $y'_i$ , 0) in the image coordinate system. However, since cameras record intensity data at discrete locations (i.e., pixels), another coordinate system named sensor system ( $x_s$ ,  $y_s$ ) is introduced. Therefore, based on the pinhole camera model, three elementary transformations must be performed. The first one transforms the global coordinates of a target object to the camera system coordinates; then, the coordinate is projected into the retinal plane; and finally, it is transformed into the sensor coordinate system in pixel units (Sutton et al. 2009). To sum up, the position of the point  $P$  in the global coordinate system is projected in the retinal plane and measured in the sensor coordinate system as  $p_1(x_{s1}, y_{s1})$  and  $p_2(x_{s2}, y_{s2})$ . Then by combining those two points with the intrinsic and extrinsic parameters of the two cameras, a 3D coordinate of the physical point P can be obtained using the triangulation theory (Luo et al. 1993).

**Fig. 1** Schematic of stereo imaging setup for 3D-DIC: reference planes and pinhole camera model transformations (Sabato and Nieszrecki 2017a)



To perform a DIC measurement, a stochastic pattern (e.g., black dots on a white background or white dots on a black background) must be applied to the surface of interest, and the relative position of each of them is tracked as the surface deforms over time. Each image can be considered as a matrix of natural integers where white pixels have a grayscale level equal to 0 and black pixels grayscale level 100. Since a single value is not a unique signature point, a neighborhood of pixels is used (i.e., facets or subsets). Each facet is a set of distinctive correlation areas defined across the measuring region. These facets are typically squares with sides of 10–50 pixels; thus, a facet will typically include several dots (e.g., gray-level variations) of the pattern applied to the structure. An example of a structural deformation and facet is shown in Fig. 2. The center of each facet can be considered a measurement point. The position of these facets spanning the field of view is tracked through each of the successively acquired images, and the 3D coordinates of the entire area of interest are calculated at each individual facet. The fundamental principle of DIC is to match the same physical point between a reference image and several deformed stages based on grayscale variations of continuous patterns. This process is shown in Fig. 2 for a more straightforward understanding. Here, the square facet with unique features is used as a reference image, while a correlation function is used for searching the corresponding facet and defining a displacement vector  $d$  as the specimen moves or deforms (Chu et al. 1985; Kahn-Jetter and Chu 1990).

Similar to the 3D-DIC, the 3DPT technique measures the 3D displacement of discrete points by tracking optical targets or fiducial markers that are placed on the structure. The 3DPT method is not a full-field measurement technique because it measures 3D motion at discrete points; however, for all practical purposes, it can be considered full field because it is relatively easy and inexpensive to mount numerous optical targets to cover the structure of interest. The centers of those markers are typically found using an ellipse finding algorithm, and the coordinates of the points in space are identified using a triangulation technique (Luo et al. 1993).



**Fig. 2** Gray-level variations on a structure with a facet and coordinate system: (a) reference stage, representing the original position; (b) deformed stage and displacement vector shown in relation to the original position. The structure (facet) has translated horizontally and vertically and has rotated (Chu et al. 1985; Kahn-Jetter and Chu 1990)

Before performing stereophotogrammetry measurements, the position of the cameras relative to each other and the distortions of the individual lenses must be determined. To obtain this information, a calibration process needs to be performed by taking several images of an object containing optical targets whose positions are previously well known (Schmidt et al. 2003a, b). Calibration is performed on the cameras' useful measurement volume to obtain the radial distortion coefficient together with the extrinsic and intrinsic parameters for each vision system. The most straightforward technique used for calibration purposes only requires the camera(s) to observe a planar pattern shown in at least two different orientations (Zhang 2000). Calibration for a field of view up to ~2 meters is typically performed by taking several pictures of the National Institute of Standards and Technology (NIST) traceable calibration objects (e.g., panels or crosses) containing optical targets (i.e., dots) whose positions are previously well known. This information is required for computing the three elementary transformations needed for the pinhole camera model to obtain the 3D coordinate of any physical point using the triangulation theory. Then a photogrammetry process known as bundle adjustment is used to establish the precise relationship between the two cameras. Once a system is calibrated, the relative position of the cameras must not be altered. Otherwise, measurement errors will occur. If the dimensions of the targeted object increase, a more complex calibration procedure needs to be performed (i.e., large-area calibration) (Poozesh et al. 2017). Once calibration is performed, camera pair image measurements can be made on a structure of interest. The simultaneously recorder pair of images can be taken at a rate that is specified by the end user, which can be on the order of milliseconds, minutes, hours, days, or even years apart. Therefore, it is possible to monitor how a structure changes over time enabling the use of 3D-DIC for SHM and NDE.

---

## Nondestructive Evaluation and Structural Health Monitoring Using Digital Image Correlation

Photogrammetry techniques have been investigated for assessing the structural integrity and the condition of engineering systems. They have proven to be practical approaches that can be used to evaluate the state of structural health and identify damage before in-service failures occur. In this section, a description of the principal NDE and SHM applications of DIC is presented. The chapter provides a broad summary of the work done by numerous researchers but is not meant to be an exhaustive list. Research efforts performed in the last few decades are divided based on their domain of use (i.e., aerospace, civil, energy, materials, vibrations, and other) for a better classification.

### NDE and SHM Using DIC in Aerospace Applications

Structural health monitoring and nondestructive evaluation of aerospace structures are critical due to the need to ensure safety. In this section, a description of the applications of DIC for monitoring the dynamics of aerospace structures and for

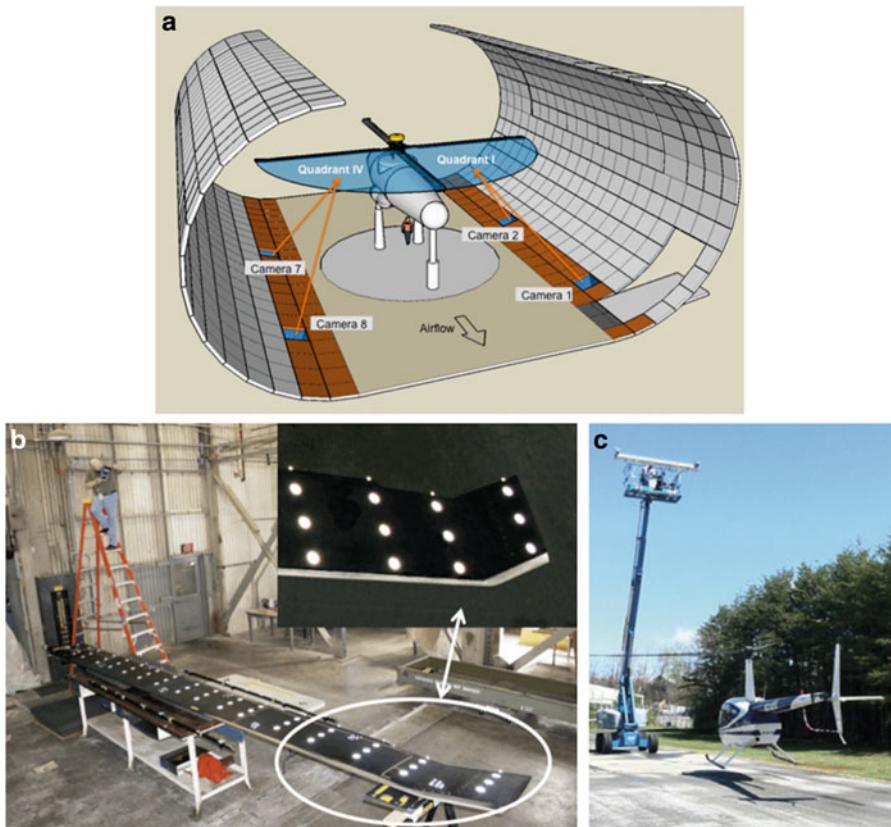
damage detection in them is presented. Then, an overview of the application for measurements on rotors is summarized.

The DIC technique was used by NASA to evaluate the structure of a helicopter in a crash (NASA 2009). A similar measurement was performed later by NASA on another helicopter (NASA 2017). One of the attempts in detecting the damage retrieved by a helicopter fuselage as a result of different loading conditions and dynamic maneuvers using 3D-DIC and dynamic photogrammetry (DP) was performed by LeBlanc et al. (2010). The test was performed on a 1:8 scale replica of an Agusta A109 helicopter and demonstrated in the laboratory the ability of DIC to measure changes caused by an added mass in the fuselage area of the aircraft, damages induced by a hard landing, tail loading, and a cracked vertical stabilizer. 2D-DIC combined with the multi-point over-deterministic method was used for characterizing the strain field around a crack propagating in a commercial aircraft wing-skin test panel during a fracture test (Du et al. 2011). The analysis of the more than 800 images both confirmed the capabilities of the optical-based technique and provided new insights into the fracture behavior of aerospace components with stiffeners and access holes. Photogrammetry proved capable of tracking the changes in the direction of the crack propagation, its width, tip velocity (i.e.,  $0.014 \text{ mm}\cdot\text{s}^{-1}$ ), and variation of the mode I stress intensity factor from measured displacement fields as increasing monotonic loadings were applied. DIC was also used to validate the assumption that reinforcement systems embedded in the panel provided very high resistance to crack growth.

DIC is a robust technique for measuring dynamics of rotating structures because it is noncontact and does not need wiring. Thus, it has been used by many researchers to monitor the dynamics of rotorcraft blades.

In 2005, a four-camera stereophotogrammetry system was used to collect flapwise, edgewise, and torsional data on a 4-meter, four-bladed rotor system operated within a low-speed wind tunnel (Schneider 2005). The work also describes efforts to accurately determine the center of rotation and transform the coordinate system from the wind tunnel coordinate system to a more convenient rotor coordinate system. In 2010, researchers from NASA performed a similar DIC test on a UH-60A four-bladed rotor system manufactured by Sikorsky Aircraft Corporation (Olson et al. 2010) (see Fig. 3a). Two-inch diameter retroreflective targets were used for image acquisition and shape extraction (see Fig. 3b) on a helicopter rotor blade. Five operating shapes of the blades were determined in the test.

In 2012, stereoscopic DIC was used in a laboratory environment to obtain the deformation of reduced-scale rotating helicopter blades with the rotor diameter of 24 and 39 inches, respectively (Sirohi and Lawson 2012). In this study, the images were captured using a pair of 2-megapixel cameras recording with a 29.5 Hz operation speed. A commercial DIC software was used to extract the bending and torsion shapes of the blades rotating at 1500 and 1800 RPM and five different pitch angles. Lundstrom et al. used the stereophotogrammetry and 3DPT to monitor dynamics of a Robinson R44 helicopter rotor and to extract its operating deflection shapes (ODS) (Lundstrom et al. 2013, 2015). In this test, the operating data from the blade of the helicopter ( $\sim 10 \text{ m}$  diameter with a tip speed of  $215 \text{ m}\cdot\text{s}^{-1}$  and a rotational frequency of 6.8 Hz) were captured using high-speed cameras as the aircraft was hovering (see Fig. 3c). The



**Fig. 3** (a) Test section schematic showing camera locations and helicopter tested in the study by Olson et al. 2010; (b) detail of the blade with 2-inch diameter retroreflective targets (Olson et al. 2010); (c) 3D-DIC measurement of a hovering Robinson R44 helicopter (Lundstrom et al. 2016)

research demonstrates that the stereophotogrammetry technique could be used to monitor the flapwise dynamics of the rotor and extract its harmonic and nonharmonic operational deflection shapes using the power spectra calculated from the acceleration (Lundstrom et al. 2011). Another attempt to measure dynamics of blades was performed when cameras were mounted to rotate with the blades. These cameras have their own batteries and can be controlled using a wireless system (Stasicki and Boden 2009). It has been shown that the operating data and ODS obtained using DIC can be used to monitor the structural motion of rotating helicopter rotors.

## NDE and SHM Using DIC in Civil Applications

Because of their large size, civil engineering structures are challenging to be monitored using wire-based sensors. Optical systems offer an alternative to the

conventional sensors because they allow for performing measurements without interfering with structure functionalities and can reach areas difficult to access without specialized equipment (e.g., trucks, scissor-lifts, and scaffoldings). In this section, an overview of DIC applications for civil structures is presented.

## Bridges

Bridges are among the most monitored civil structures due to their strategic importance and safety issues. The use of computer vision methods to study the dynamic characteristics of bridges and to assess their health status goes back to 1999. Olaszek proposed to use photogrammetric principles for real-time monitoring of the displacement for a point on a bridge. By using a charge-coupled device (CCD) camera with a telephoto lens, he measured the displacement of a target (i.e., a black cross on a white background). His vision system was equipped with an additional reference to decrease the sensitivity to vibrations (Olaszek 1999). One of the first studies performed to apply optically based techniques for damage detection and SHM was the work by Fu and Moosa in 2002. In their research, the authors compared the results of measurement performed with CCD cameras with dial gages and laser devices to prove that cameras could be used for measuring structural displacement (Fu and Moosa 2002). Analogously, Lee and Shinozuka, in 2006, proposed and validated a target recognition algorithm to measure the dynamic displacement of bridges in real time using images. They validated the technique using shaking table tests and an open-box girders bridge. The laboratory tests showed that the proposed system was able to measure displacements with an accuracy less than 3% error in maximum values when the measurement was compared to those recorded using a linear variable differential transformer. In the field applications, the displacement measured by the proposed system showed close results to the laser vibrometer used for a back-to-back comparison with small measurement noise. The main difference of this work to the previous studies was the use of consumer-grade cameras with no specially manufactured optical devices and the absence of an off-line sophisticated signal processing algorithm (Lee and Shinozuka 2006). Following that, in 2007, Yoneyama et al. used DIC to monitor the variation of vertical deflection of bridge girders. In their research, a bridge was loaded by a heavy cargo truck, and the images of the bridge girder surface were recorded by a single digital camera before and after deformation. The results showed that deflections obtained by DIC agree well with those recorded by transducers. Moreover, results from other tests performed showed that the bridge's deflection could be measured even when the random pattern was not applied, and only features on the surface of the tested object were used (Yoneyama et al. 2005; Yoneyama et al. 2007). In 2009, Zaurin and Catbas integrated video images and sensor data to obtain unit influence line (UIL) as an index to monitor bridge behavior under identified loading conditions. Although this work is not entirely related to DIC, it showed another aspect of using vision-based methods by integrating computer vision techniques and sensor data to gain information for damage detection and condition assessment for decision making (Catbas et al. 2011; Zaurin and Catbas 2009, 2011). In 2010, Malesa et al. compared the performance of DIC with a FEM to measure the displacement of a railway bridge.

To reduce the computational power of the DIC algorithm, the researchers implemented a custom-built software package that also considered a normalized correlation metric for outdoor measurements under changing lighting conditions. During the tests performed to validate the proposed methodology, the displacement of optical markers was recorded with a single 1-megapixel camera as a locomotive was passing at different speeds through the bridge (Malesa et al. 2010). In 2011, Chiang et al. developed two DIC computer codes (i.e., DIC-c and DIC-s) and validated their performance by measuring the vertical displacement of a three-span highway bridge by using three individual marked plates, spaced at 5 meters when a full-size truck was passing by the measurement point. In the tests, a user-grade camera installed with an AF18–270 mm lens located 50 meters away from the bridge was used. Results showed that the proposed system was computationally demanding since it was capable of obtaining the displacement field of a 1-megapixel image within an hour at a resolution of 1/8X1/8 pixel (Chiang et al. 2011). In the same year, researchers at the University of New Hampshire conducted laboratory and field experiments to validate the possibility of using DIC for measuring the vertical displacement of a three-span continuous steel girder bridge and a short concrete slab culvert with a fiber-reinforced polymer retrofit. A comparison between the results from DIC and LVDT sensors showed that DIC results were within 0.03 mm of LVDTs. Also, DIC data was used for model verification in the study (Peddle et al. 2011). In 2012, Busca et al. proposed to use a Canon consumer-grade camera equipped with a  $1920 \times 1080$  pixel sensor and three different levels of zooms to measure the displacement of a 50-meter-long steel truss bridge crossing a river as trains traveled at low speed over it. During these tests, optically retrieved data was compared with those obtained from a single-point laser interferometer scanner and a wire potentiometer placed in the mid-span of the bridge. The camera was located nearly 100 meters away from the target object, and three resolution levels were used for tracking the displacement of optical targets deployed on the bridge's trusses. In this test, the higher level of zoom was capable of producing more accurate results (i.e., 0.3 mm/pixel); however, it created a smaller FOV, which was of less interest as the full-field response of the structure needed to be analyzed. A strong agreement was shown between the data recorded using the DIC method and the laser even with a low level of excitations (e.g., after the train's passage). Moreover, the optical system was able to detect the first vibration mode of the bridge around 5 Hz. This research also showed the robustness of the method in the case in which no optical targets were deployed on the test structure. In this case, researchers worked on the image contrast between the main beams and the sky. Despite the fact that this approach provided less accurate results, the oscillatory behavior of the bridge was still captured in the measurement (Busca et al. 2012). Moreover, it was shown that measurement reliability is strongly improved by the structure texture contrast. The same group of researchers also proposed a comparison between three different image processing algorithms (i.e., edge detection, pattern matching, and digital image correlation) to estimate the vertical deflection of the same railway bridge subjected to train pass-by (Busca et al. 2014). In 2013, Ye et al. developed a vision-based dynamic displacement measurement system with digital image processing

technologies. The researchers used a high-resolution industrial CCD digital camera and an extended-range zoom lens capable of capturing the digital images of a target on the structure over 1000 meters away to trace and identify the displacement of a targeted structure with the aid of pattern matching algorithms. The proposed system was validated with both laboratory tests using well-known excitations and an in-situ experiment to measure the mid-span vertical displacement of the suspension Tsing Ma Bridge (see Fig. 4a) and the cable-stayed Stonecutters Bridge during operational conditions. The developed system showed robust capability of long-distance remote displacement measurement and a good agreement between the optically based



**Fig. 4** (a) Vision-based displacement measurement of Tsing Ma Bridge (Ye et al. 2013); (b) prototype UAV with 3D-DIC payload; (c) observed crack with in-plane full-field displacement contour (Reagan et al. 2017b); (d) precision optical target and lighting lamps used as part of the advanced video-based system used for monitoring the dynamic displacement of a railway bridge (Ribeiro et al. 2014); (e) railway bridge dynamic displacement measurement under trainloads (Feng et al. 2015a)

measured vertical displacements and those measured using a GPS system located on the first bridge. In addition, the vision system was able to measure the vertical displacement influence lines of the bridges under different loading conditions (Ye et al. 2013). In the same year, Kohut et al. compared the performance of DIC and radar interferometry (RI) in measuring the deflection of civil structures. They compared the displacement field of a steel viaduct subjected to an operational load due to tram traffic. While the RI system used a set of reflecting elements for performing the measurement, the DIC uses a single camera located ~28 meter away from the bridge to record the position of a plate with a speckle pattern on it. In order to obtain a uniform scale in the entire field of view, the lens axis was aimed at a perpendicular direction to the span axis. Performed tests proved a better stability and robustness to noise for the vision-based system because of, probably, the uncertainty of the points reflecting the RI wave. Moreover, it was noted that the RI was a 1D system, while a single camera was able to provide 2D information (Kohut et al. 2013). In the same year, Koltsida et al. demonstrated the use of DIC under laboratory conditions for small-scale masonry specimens and field conditions for a masonry arch railway bridge under traffic loading. Because of the inherent features of the masonry, an artificial speckle was not required during the tests, as the natural texture of the material's surface has random gray intensity distribution. In these experiments, DIC was used for measuring strain, cracks, and crack openings under increasing loads. Moreover, a pilot test was performed on a four-span masonry arch railway bridge to measure horizontal and vertical displacements and visualize deformation across the surface as trains were passing by (Koltsida et al. 2013). Nonis et al. used 3D-DIC for periodic inspection of concrete bridges to locate nonvisible cracks in concrete, quantify spalling, and measure bridge deformation. In a laboratory test, they demonstrated that optical-based measurement correlated well with those performed using fiber-optic strain gages during three- and four-point bending tests conducted on a concrete beam. Then, they used photogrammetric targets as extensometers to track the opening of joints and cracks over a 4.5-month period and a stochastic pattern to monitor the strain fields over two bridges in service. In the same study, they also used a projected pattern to quantify spalling phenomena (Nonis et al. 2013). The same bridges were monitored for almost one year using 3D-DIC by (Reagan et al. 2017a). The researchers proposed a novel approach that combines the use of an unmanned aerial vehicle (UAV) and 3D-DIC to perform noncontact, optically based measurements to monitor the health of bridges (see Fig. 4b). By installing a stereovision camera system on a drone payload, extensive laboratory tests, and long-term monitoring campaigns, they demonstrated the accuracy of this system in detecting structural changes and monitoring the dynamic behavior of hairline cracks and expansion joints over time. Results show that the combined 3D-DIC UAV system was able to perform structural investigations and detect changes to the bridge geometry with an uncertainty on the order of  $10^{-5}$  m. These results outperformed the resolution that can be obtained when visual inspection techniques are employed while improving accessibility (Reagan et al. 2016, 2017b).

In 2014, Ribeiro et al. developed a noncontact displacement system for railway bridges monitoring based on a high-speed video camera (from 64 fps to 500 fps), an optical lens, lighting lamps, and a precision target. The system was used to measure

the displacement of a railway bridge's deck, induced by the passage of trains at speeds up to 180 km/h. Results showed that system precision was below 0.1 mm for a distance of the camera to the target up to 15 meters and nearly 0.25 mm for a range of 25 meters when compared with LVDT data. Findings of their study also proved that the system was sensitive to camera movements, and heat waves generated by IR incandescent lighting increased the FOV distortion (Ribeiro et al. 2014). In 2015 Feng et al. developed an advanced template matching algorithm (i.e., upsampled cross-correlation) for real-time displacement extraction from video images with sub-pixel accuracy using Fourier transform. Because of the upsampling cross-correlation method adopted, the system was capable of measuring structural vibrations smaller than 1 mm by adjusting the upsampling factor. The system was validated through extensive laboratory shaking table tests, by measuring the displacement of a high-contrast artificial target attached to the test structure, and in field tests. In this work, the vertical deflection of a railway bridge during train passing was measured, and the frequency domain response of a pedestrian bridge subjected to dynamic loading was used for validating the performance of the proposed system (Feng et al. 2015a). Data recorded using the vision system were also used to validate a finite element model (FEM) of the same railway bridge during the passage of different trains (Feng and Feng 2015). Moreover, the performance of the system was evaluated when low-contrast natural targets or structural features were used as optical markers (Feng et al. 2015b). In 2016, Pan et al. advanced a video deflectometer system using off-axis DIC for real-time measurement of vertical bridge deflection. They used an inverse compositional Gauss-Newton algorithm capable of detecting image displacements in pixels and converted them to physical movements using a calibration technique with the aid of a laser rangefinder. This system, similar to the one proposed by Feng et al. 2015a, is capable of detecting in-plane displacement only because of the adopted pinhole camera model. To demonstrate the accuracy of the system, field measurements of the deflection of a 60-meter three-span railway bridge were performed. The vision-based system proved to be able to measure the dynamic displacement of the bridges, but results were noisy when compared to traditional measurement systems, confirming its sensitivity to various external disturbances (Pan et al. 2016).

## Hanger Cables

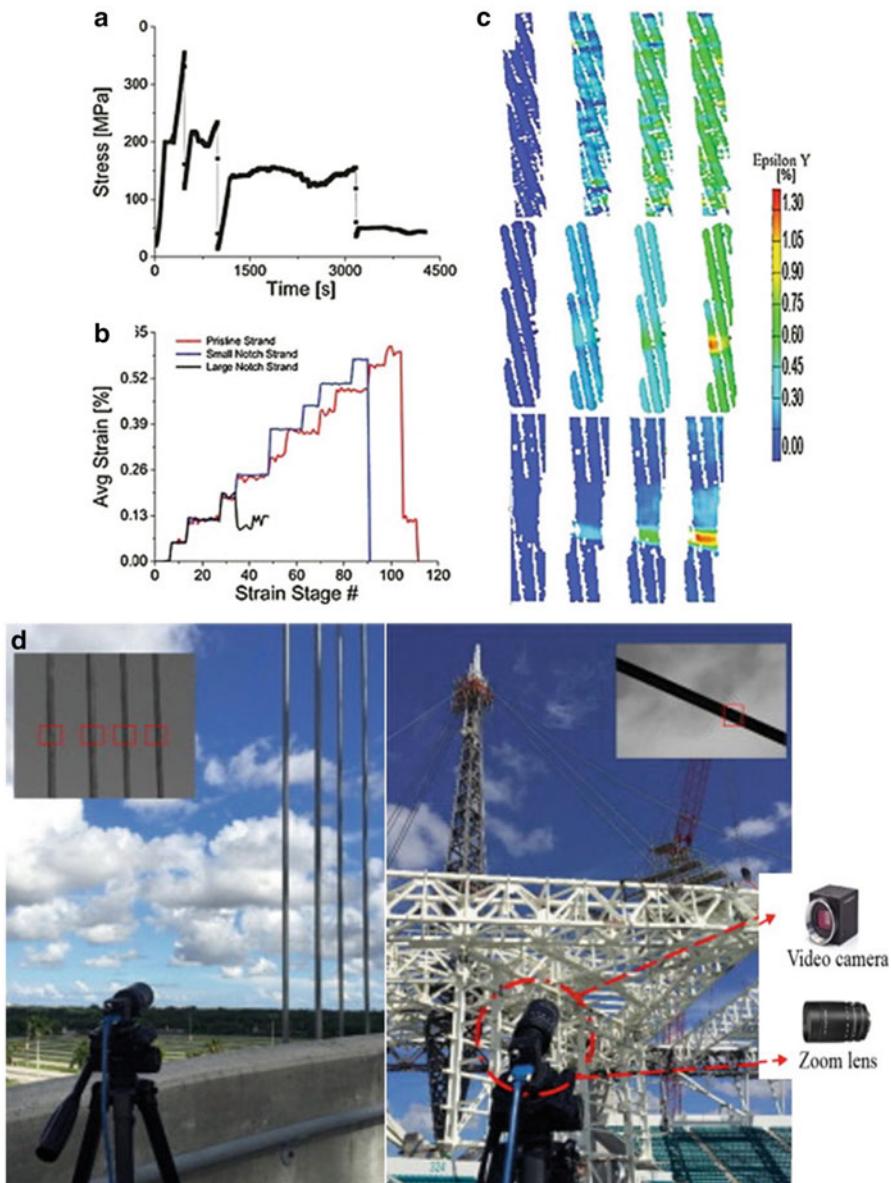
Another popular application of optically based measurement for civil structures and infrastructures is the monitoring of bridge hanger cables. In 2008, Ji and Chang were among the first to propose to detect the vibration of the wires holding a small pedestrian bridge using vision systems. By using a commercially available camera, the proposed method analyzed a sequence of images of a vibrating cable and calculated the variation of optical intensity of an arbitrarily selected region of interest to measure the displacement of the wire (Ji and Chang 2008). In 2013, Kim and Kim used DIC to develop an image processing algorithm to analyze images acquired through a digital video camera to estimate tension in the Gwangan Bridge's hanger cables by measuring its dynamic response. The measurement was performed using only the cable shape and without the aid of optical targets deployed on the structure.

In the study, the motion of the vision-based system was also corrected considering a fixed object in the images as a reference point. The results obtained using digital image processing methods and those measured using accelerometer-based methods were characterized by a difference of  $\pm 0.5\%$  for both modal frequencies and tension (Kim and Kim 2013).

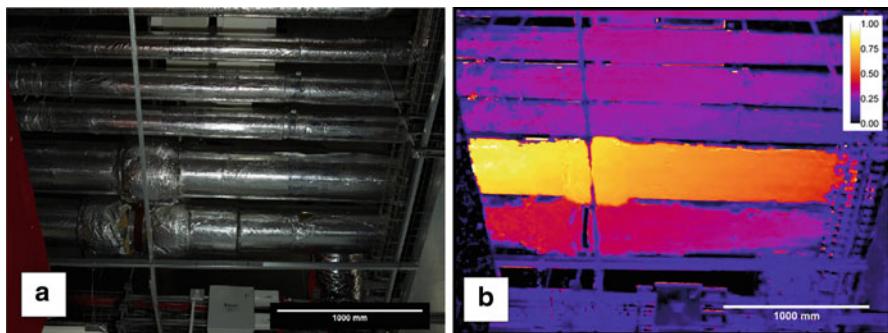
In 2013, Vanniamparambil et al. proposed a sensing system combining 3D-DIC, guided ultrasonic waves (GUW), and acoustic emission (AE) to detect breaks on seven-wire steel strands. 3D-DIC was used to identify full-field strain accumulations on the surface of the strands and the notched areas before cable failure (see Fig. 5a–c). The results showed that the damage source location could be identified by fusing strain data with acoustics measurement to perform NDE analyses on pre-stressed and post-tensioned cables found in stay cable bridges (Vanniamparambil et al. 2013). In 2016, Ye et al. employed the multi-point pattern-matching correlation algorithm (Ye et al. 2013) to improve a cable force monitoring system. The accuracy of the vision-based approach was validated by conducting uniaxial tensile tests of steel bars, steel wire ropes, parallel strand cables as well as a series of moving loading experiments on a scale arch bridge model. Laboratory tests proved that results obtained with the vision-based system agreed well with those obtained using traditional force sensors (Ye et al. 2016a). In 2017, Feng et al. used a vision-based sensor system, based on the sub-pixel orientation code matching algorithm (Feng and Feng 2016) to measure the forces acting on the cable-supported roof structure of the Hard Rock Stadium in Florida and ensure that cable tension reach their design values (see Fig. 5d). In their tests, the researchers used a 2-megapixel CMOS-type camera run with a sampling rate of 50 fps. The measured displacement of the cables was used to calculate the natural frequency and evaluate the cable tension. Analyses of the results showed that the measurement recorded utilizing the vision system was useful in determining the force acting in the cable with a maximum error below 5% when compared with reference accelerometers (Feng et al. 2017).

## Other Civil Applications

3D-DIC has also been employed for several other applications in the field of civil engineering. For instance, in 2005, McGinnis et al. investigated the use of DIC to measure displacement and strain fields during the use of the hole drilling method to evaluate the state of stress in concrete (McGinnis et al. 2005). The same research group developed an experimental set-up protocol for deploying multiple 3D-DIC sensors simultaneously during the same structural test. The experiments were designed to allow four camera pairs to capture the deformations of key structural components of a coupled post-tensioned shear wall system (McGinnis et al. 2012). McCormick and Lord in 2012 started to explore this technique for monitoring the vibration of cooling water pipelines (see Fig. 6) or the movement of the London Eye. What is interesting in this study was the capability of the vision system to detect displacement relying on the features of the object being tested



**Fig. 5** Bridge cable (a) stress and average strain curves; (b) full-field strain for the large notch strand; (c) strain visualizations for pristine (top), small notch (center), and large notch strand (bottom) measured using 3D-DIC during a load-unload test performed on a seven-wire steel strand (Vanniamparambil et al. 2013); (d) vision-based system for determining the tension force in the cable-supported roof structure of the Hard Rock Stadium in Florida (Feng et al. 2017)



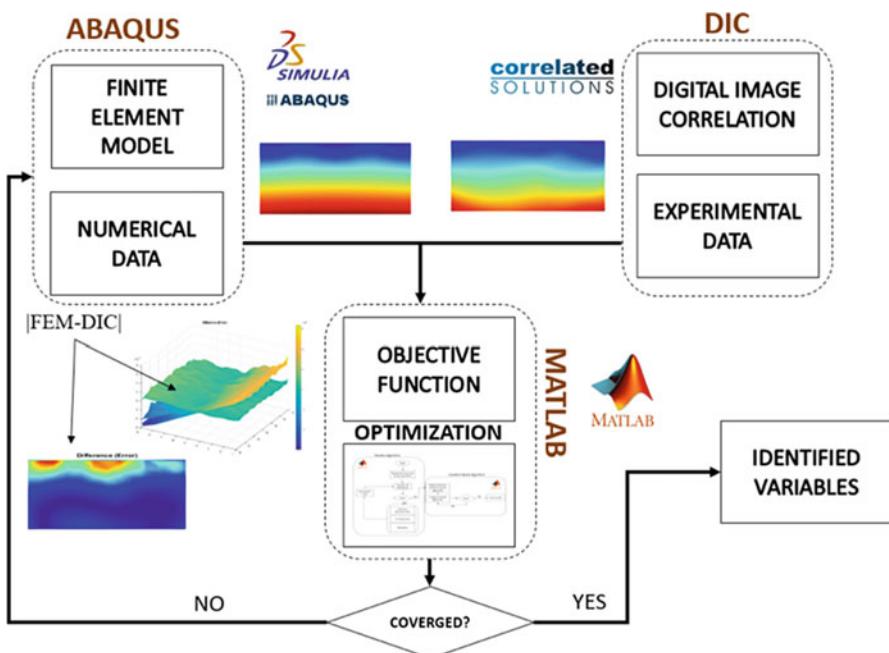
**Fig. 6** 3D-DIC used to detect the out-of-plane displacement of water cooling pipelines (McCormick and Lord 2012)

(e.g., the thermal insulation covering of the pipes) (McCormick and Lord 2012). In 2014, Kohut et al. proposed a vision-based method based on DIC to measure deformations of civil engineering structures under loading condition. The innovation of the proposed approach consisted of three steps: (i) image rectification, (ii) displacement field measurement, and (iii) scaling (Kohut et al. 2017). The first step of the method is optional and is performed whether the images are taken from two distinct points in space. In the second step, the image is divided into intensity patterns. Corresponding patterns are identified on the reference image through the correlation algorithm. The deflection curve is computed as the difference between positions of the similar image patches on two images. Finally, the scale coefficient is calculated from objects with known geometric dimensions. The authors validated their findings by measuring the deflection of a crane under load and comparing the results with those obtained using a laser tracker (Kohut et al. 2014a, b; Sładek et al. 2013).

In 2014, Murray et al. used four synchronized, 85 mm focal length, high-speed cameras to measure the out-of-track motion of the vertical and longitudinal rail displacement using the texture on the rail itself and tie-mounted targets. The cameras were mounted at a fixed position near the track and away from the train. The proposed system was evaluated at two different sites, one with a high-quality subgrade and one with a peat subgrade, during the passage of 11 trains at the monitored location. Data recorded using the four cameras were used to determine the stiffness and damping of the subgrade and to investigate the factors that influence the magnitude of longitudinal rail displacement (Murray et al. 2014). Sabato and Niezrecki in 2017 proposed another railroad track inspection technique using 3D-DIC. They investigated the feasibility of using optical systems deployed underneath a rail car to assess the deformation profile of railroad crossties for different ballast conditions and under increasing loads. Measurements to validate the proposed approach were conducted using both a painted pattern and a stochastic pattern projected on the crosstie using a projector. Measurements made while using a projected pattern were shown to be equivalent to those performed as the speckles were painted on the target object. Extensive series of laboratory tests showed that the

3D-DIC system was able to detect out-of-plane displacement as low as  $10^{-4}$  m and with an error of about 3% when compared with a traditional wired, high-accuracy LVDT (Sabato and Nieszrecki 2017a, b). The researchers also investigated the effect of train-induced vibration on the cameras, demonstrating that an isolation system tuned at about 2 Hz was effective in significantly reducing the vibration transmitted to the vision sensors (Sabato et al. 2017b).

Researchers from the University of Virginia proposed studies to perform an analytical-computational correlation between DIC and FEM. Despite that their research was mainly focused on structural civil engineering applications, their findings can be easily applied to other fields. Dizaji et al. proposed to use 3D-DIC in conjunction with a FEM approach interfaced with MATLAB for structural identification on a large-scale specimen (i.e., concrete beam under different loading and boundary conditions). Full-field results, obtained through the optically based method, were validated against ground-truth measurements from traditional sensors and used to update the FEM via a hybrid genetic algorithm that combines a genetic algorithm to perform global optimization in conjunction with a gradient-based method as summarized in Fig. 7. In their research, the authors used both dynamic and static responses obtained from vision-based measurements to minimize the matching errors between 3D-DIC and FEM, proving that this approach was effective for large-scale structural systems (Dizaji et al. 2017a, b). The authors also validated



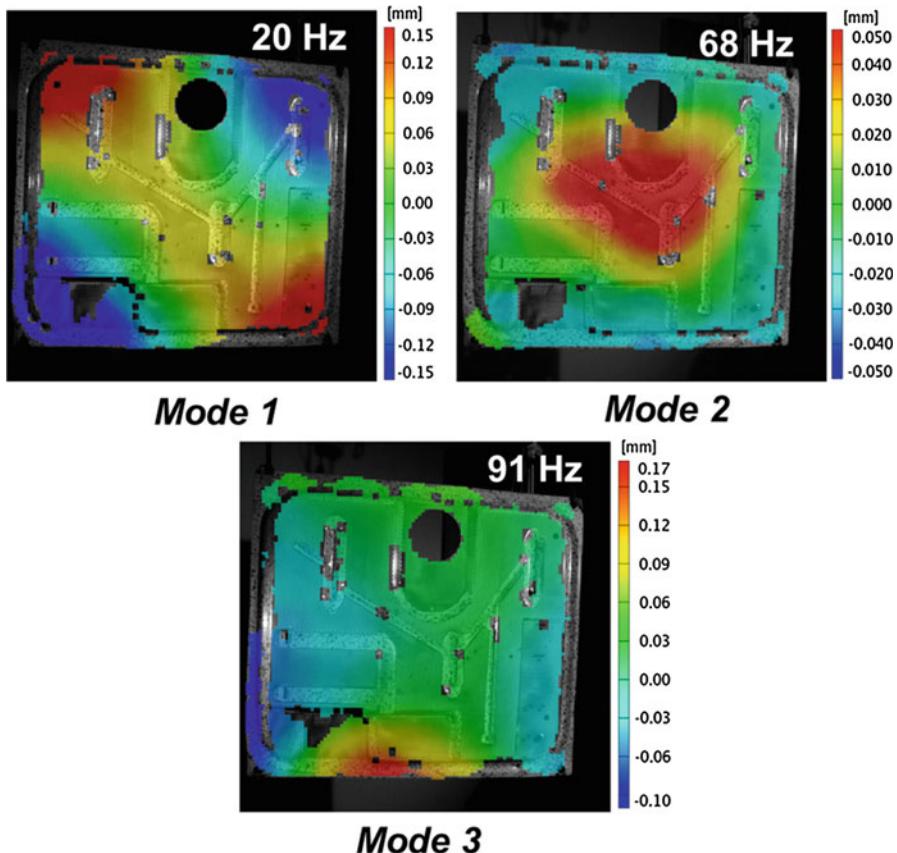
**Fig. 7** 3D-DIC FEM integrated approach for structural identification of large-scale structures (Dizaji et al. 2017a)

the use of DIC to characterize the behavior of concrete across different structural scales by performing a number of experiments (i.e., compression, split tensile, and flexural tests) at both mixture and structural member scale levels.

The work evaluated the performance of two representative plasticity-based numerical models commonly used in stressed concrete characterization (Gheitasi et al. 2018).

## NDE and SHM Using DIC in Dynamic Applications

Optical systems are robust alternatives to traditional accelerometers for performing structural dynamic analyses and determining modal parameters of interest in SHM and NDE applications (e.g., mode shapes and structural frequency response). One of the first attempts that used digital image processing for system identification and determining nonlinear characteristics of mechanical and structural systems was performed by Chung et al. in 2003. The study demonstrated the proof-of-concept that optical techniques can be used to identify structural characteristics by digitally processing the analog image captured using a videotape recorded during a shaking table test of a model structure (Chung et al. 2004). Following that, Helfrick et al. used 3D-DIC to perform full-full field vibration measurements for damage detection (Helfrick and Nieszrecki 2007). In their study, the researchers used 3D-DIC techniques to detect crack across the width of a cast acrylic cantilever beam whose depth was equal to 17% and 40% of the beam thickness. During the tests, the structure was excited using an electro-mechanical shaker at each of its first three natural frequencies corresponding to the first three bending modes of the structure. DIC data was used to calculate the curvature of the beam and polynomial best-fit curves for each damage condition and each excitation frequency. The analysis of the results showed that the optical system was able to detect the presence and location of the damage for the 40% thickness-reduction case using the second and third natural frequencies excitations. For a 17% thickness reduction, the damage index plot derived from DIC measurements contained mostly the same degree of fluctuations as the undamaged condition and was not useful in providing information about the damage (Helfrick et al. 2009a). The same authors started to employ 3D-DIC for full-field vibration measurement. They used optically based measurements for modal analysis to measure the shape and deformation of a vibrating metal base of a dryer cabinet (Helfrick et al. 2008). The study presented results that compared the DIC-measured data with similar tests performed using traditional devices (i.e., accelerometers and a scanning laser Doppler vibrometer) and also a finite element model. The study proved that the displacement measured using a pair of synchronized 4-megapixel high-speed cameras had a noise floor of  $\sim 20 \mu\text{m}$  when the measurement volume was  $75 \times 710 \times 710 \text{ mm}^3$ . DIC data correlated very well with the other measurement techniques (i.e., accelerometer and laser vibrometry), and the 3D-DIC measurement was able to extract the first three mode shapes of the structure (see Fig. 8). The number of points extracted by each approach was significantly different. The modal hammer test extracted information at about 30 points, the laser 313 points, and the DIC



**Fig. 8** Contours of out-of-plane displacement overlaid on a dryer panel using 3D-DIC (Helfrick et al. 2011)

measurement had ~7000 points highlighting the high spatial density of the DIC measurement. What was also impressive was the time required for performing the data acquisition. In particular, DIC was able to provide results from images recorded for less than one second, while the study reports that the laser measurements took nearly 10 hours to complete, highlighting one of the most significant advantages of optical techniques compared to traditional approaches (Helfrick et al. 2011).

To show the benefits of using optically based methods for performing high-displacement and low-frequency vibrations measurement, typically difficult to measure with accelerometers and laser vibrometers, Warren et al. conducted a back-to-back comparison between 3D-DIC, 3DPPT, 3D laser vibrometry, and accelerometers to measure the dynamics of a structure. In the tests, 3D-DIC and 3DPPT monitored the response of a base-upright steel structure at the same eight locations as performed for the accelerometer and laser tests. While DIC can be used over the whole surface, in this case, an alternative approach with discrete patches of patterns

was employed to illustrate that equivalent data could be extracted without having to pattern the entire structure. Because of the specification of the cameras used (i.e., 2 megapixel), working distance (2 m), noise floor ( $\sim 40 \mu\text{m}$ ), amplitude of the displacement characterizing higher frequencies, and ambient vibrations, only the first and third modes of the structure (i.e., 26 and 78 Hz, respectively) were captured. Similar tests were performed using 4-megapixel high-speed camera recording at 500 Hz. Additionally, the displacements determined using 3DPT were used to calculate frequency response functions (FRFs), from which mode shapes were extracted. The results indicate that when low-speed cameras were used in conjunction with forced normal mode testing, both 3D-DIC and 3DPT could accurately capture mode shapes as long as measurable displacements were present. The high-speed optical results were the best obtained in these studies, and the first five modes of the structure (i.e., up to 108.8 Hz) were determined each having modal assurance criterion (MAC) values better than 99% (Warren et al. 2011b, c).

In 2011, researchers started to use DIC algorithms to estimate the vertical motion of people in a crowd and determine the loads generated by the people jumping on the hosting structures (Caprioli et al. 2011; Mazzoleni and Zappa 2012). In the study, an area correlation matching algorithm and a minimum quadratic difference method are used to analyze two subsequent images recorded by a conventional digital camera and depicting the crowd on stadium grandstands. The study focuses on understanding the effect of different parameters on the measurement accuracy, and it was based mainly on computer simulation. The robustness of the method was experimentally verified by comparing the signal obtained from DIC analyses of a video of a group of people jumping and the signal recorded by an accelerometer attached to one of the people in the crowd. This study paved the road for another study in which the structural response of a major UK stadium was measured (Jones et al. 2011). Using DIC approaches, video of people jumping on the grandstand of the stadium recorded during major sports events was used to determine the forces induced by the crowds and calculate the loads applied to the structure. These forces were then used in simulations involving modal space approximations of a large-scale complex FEM to determine the structural response of the targeted system, and the comparisons generated MAC values all above 65%.

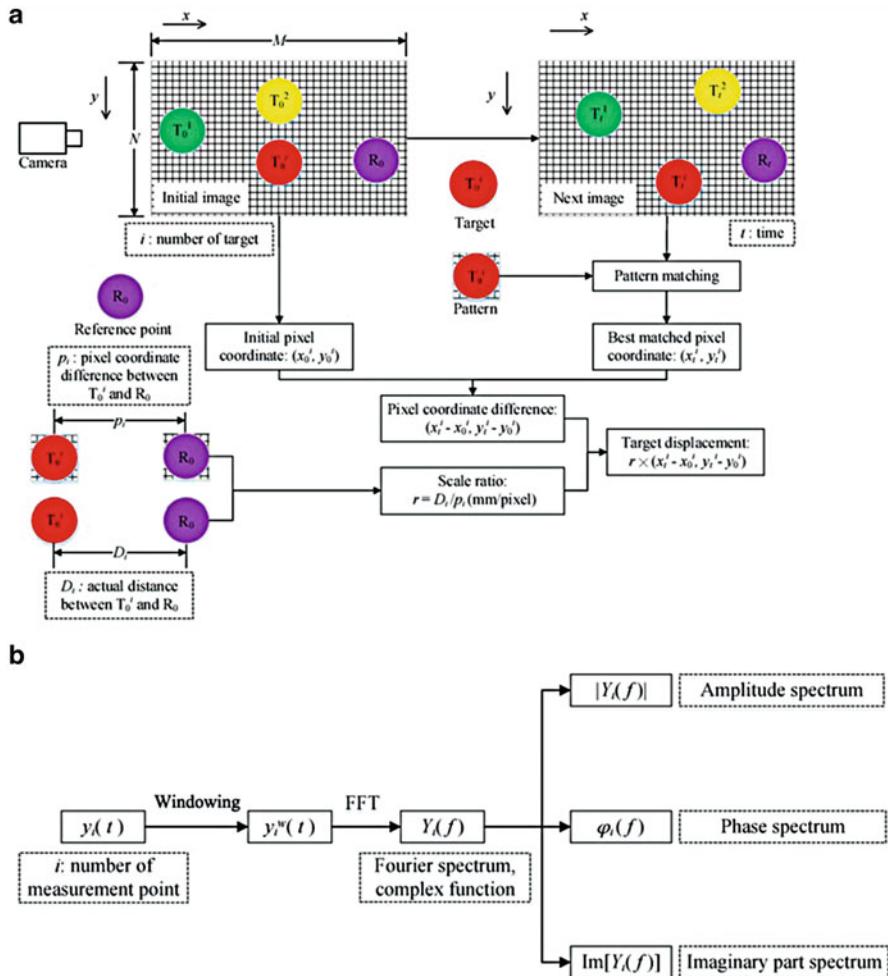
In 2014, Wu et al. developed a 2D, 60 Hz sampling rate, vision system to monitor plane vibrations of a reduced scale frame mounted on a shaking table using images recorded by a commercially available digital camera. The study focused on determining the importance of camera parameters, the trade-off between the system resolution and the FOV, and the upper limitation of marker density. In their investigation, to overcome unstable experiment precision due to the scale factor approach used to convert image coordinates measured by a vision system in the unit of pixels into space coordinates, the researchers developed two alternative methodologies: registration and direct linear transformation. The authors processed images using a customized MATLAB code to obtain the spatial coordinates of the markers. They demonstrated that motion with frequency up to 20 Hz could be captured and successfully analyzed in time and frequency domains to determine the dynamic characteristics of the targeted structure (Wu et al. 2014).

Bartilson et al. improved the findings of Cigada et al. 2013 by developing a targetless vision system for structural dynamic response studies (Bartilson et al. 2015) to reduce the difficulty associated to the deployment of targets to the test structure, especially in locations critical to access. The main difference of this study with others was the use of a minimum quadratic difference (MQD) algorithm which was more robust than the previously employed edge detection and cross-correlation algorithms. In this technique, no targets or patterns were placed on the structure of interest, and consumer-grade cameras were used. The authors validated their method by accurately determining natural frequencies of a full-scale traffic signal structure with accuracy similar to strain gages and string potentiometers. By utilizing time series filtering techniques, modal damping ratios were accurately determined, and the mode shapes of displacement reaching only 0.5 pixels were obtained. To continue making optically based measurement easy to perform, in 2017, Khuc and Catbas proposed a novel approach for displacement and vibration analysis by exploring the possibility of using a new type of virtual markers instead of physical targets. The study suggested a practical camera calibration method to calculate the converting ratio between pixel-based displacement and engineering units and analyzes best practices to consider low contrast, changing illumination, and outliers in matching key points. The method was validated on a four-span bridge model and a real-world structure with excellent results for both static and dynamic behaviors of the two structures (Khuc and Catbas 2017).

In the same year, vision-based measured displacement was used to identify structural parameters and external forces. In their work, Feng and Feng aimed to link the measured displacement data to the quantification of the structural health condition of a structure. By using output-only vision-based displacement measurement, the authors validated the feasibility of simultaneous identification of structural stiffness and unknown excitation forces in the time domain. Laboratory tests performed on a simply supported beam specimen demonstrated that two measurement points by the vision sensor were sufficient for accurately identifying both the beam stiffness and hammer impact forces (Feng and Feng 2017).

In 2018, Dong et al. proposed a method, based on the multi-point pattern matching algorithm, for simultaneous multi-point measurement for structural modal parameters identification using a FFT. A description of the algorithm can be seen in Fig. 9. In the study, comparisons with the results obtained by the vision-based system and accelerometers were performed together with analyses on the effect of the distance measurement on the accuracy of the vision-based system and the feasibility of different types of targets (i.e., LED lamps and black spots). Results of experiments performed on a simply supported rectangle steel beam model showed how the developed vision-based system could identify dynamic response and the first two modes of the targeted structure with good accuracy for a distance up to 15 meters with errors between 6 and 14% when compared with a theoretical model of the beam (Dong et al. 2018).

This section presented a description of structural dynamics potentiality of 3D-DIC and 3DPT. For a complete summary of the state of the art of full-field optical techniques specifically for structural dynamics measurement, the interested readers can refer to Baqersad et al. (2017) and Niezrecki et al. (2010).



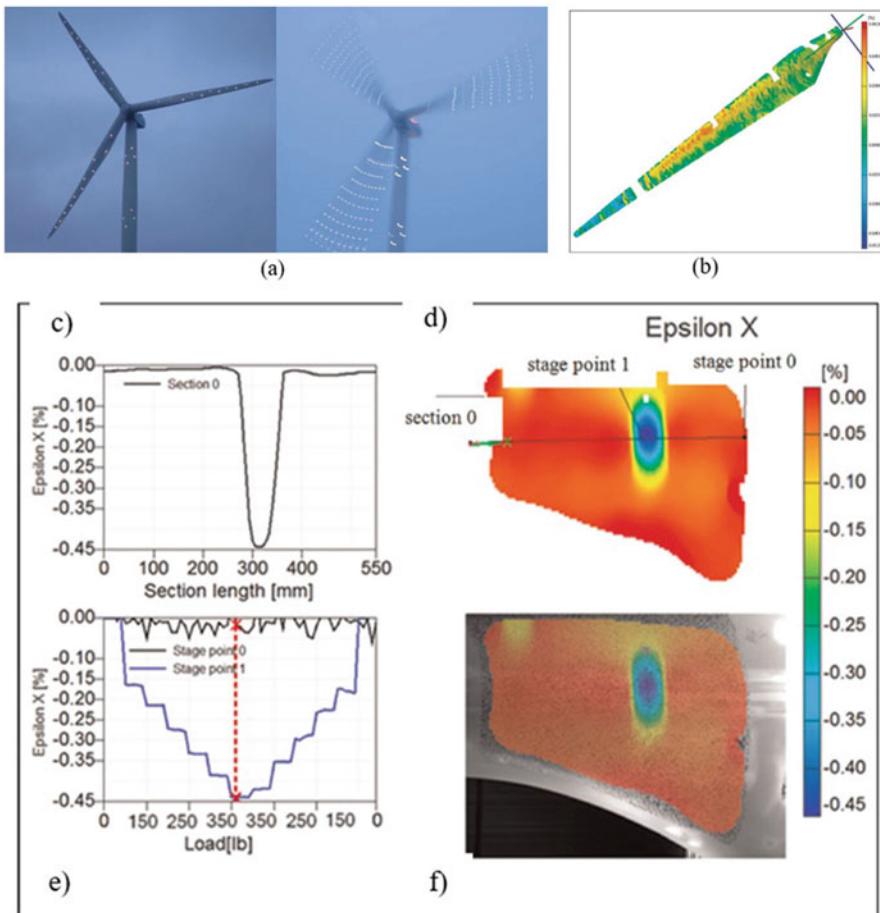
**Fig. 9** Flowchart of the multi-point pattern matching algorithm proposed by Dung et al. (2018) for displacement measurement (a) and time-frequency domain transformation (b) to perform modal parameter identification (Dong et al. 2018)

## NDE and SHM Using DIC in Energy Applications

The growing need for clean energy has spurred the widespread use of wind turbines as a valuable alternative to traditional fossil fuels. However, as the size of these machines increases to meet energy demands, wind turbines are subjected to a combination of increased static and dynamic loadings that have an impact on their performance, efficiency, and reliability. Certifying wind turbine blades is an essential part of the design process to verify structural integrity as well as fatigue life. DIC has been demonstrated to be an excellent optical-sensing technology for improved wind turbine blade SHM and NDE.

The DIC technique and its suitability to achieve the SHM and NDE goals of identifying strain amplification or excessive deformation in regions of damage have been explored in several studies. In 2009, one of the first works to use stereophotogrammetry for large wind turbine blades collected dynamic operating data to measure the dynamic behavior of a 500 kW wind turbine during an emergency stop from 24 to 0 RPM (Paulsen et al. 2009). In the same year, researchers proposed a method to use low-speed cameras to measure vibrations in rotating structures by using phase stepping (Helfrick et al. 2009b). After that, high-speed cameras were used to estimate vibrations in a small-scale rotating wind turbine, but only a few operating deflection shapes (ODSs) could be determined (Warren et al. 2011a). ODSs extracted from 3DPT data were compared to other rotating tests and a nonrotating modal test. Results indicated a strong correlation between conventional static and optically measured mode shapes. However, some spectral differences were found, and researchers assumed they could be due to changes in the structural boundary conditions present during operation at different speeds. Following this work, Ozbek et al. used retroreflective optical targets to measure the displacement of on a 2.5 MW Nordex N80 wind turbine with an 80 m height tower and rotor diameter (Ozbek et al. 2011, 2013; Ozbek and Rixen 2013) (see Fig. 10a). Within these studies, the researchers describe the efforts to sufficiently illuminate this massive structure and collect stereophotogrammetric operating data. The structure was illuminated with high-power light-emitting diode (LED) strobe lights synchronized with the camera pair by a central computer. During in-field tests, the dynamic behavior of the turbine was monitored from a measurement distance of 220 m by using a 3DPT technique, and results showed that the deformations of the turbine could be measured with an average accuracy of  $\pm 25$  mm. In the study, it was observed that the measurement error was frequency dependent and mainly caused by calibration problems due to the rotation of the turbine. It was also found that measurement accuracy improved for higher rotational frequencies (i.e., above 1.4 Hz). Also, recorded data were analyzed by using an operational modal analysis algorithm based on the least square complex exponential method, and several turbine parameters (Eigenfrequencies and damping ratios) were extracted for different wind speed.

Between 2011 and 2013, LeBlanc et al. used 3D-DIC for the full-field inspection of a 9-m-long TPI composites CX-100 wind turbine blade manufactured for Sandia National Laboratories. The goal of the study was to extract full-field displacement and strain measurements from a composite turbine blade subjected to increasing static loading (LeBlanc et al. 2011a, 2013) (see Fig. 10b). The use of 3D-DIC allowed the observing of significant strain amplification in damaged areas, as well as discontinuities in the curvature of the blade at the location of damages. The optically based technique was able to quantify the progression of damage as the load was applied during laboratory tests, providing more structural information than discrete point-strain measurements (see Fig. 10c). Also, a method for combining multiple measured fields of views (FOVs) of the same object recorded moving around the stereovision system (i.e., stitching) was proposed to obtain the deflection and strain over the full length of the 9 m wind turbine blade. It is important to point out that stitching of single geometry and displacement fields is performed by finding



**Fig. 10** (a) Retroreflective targets placed on a 2.5 MW NordexN80 wind turbine used in the study by Ozbek et al. 2013; (b) span-wise full-field strain of blade obtained by stitching together 16 independently recorded fields of view as the blade was subjected to a load of 350 lb. located 6.75 m away from its root section (LeBlanc et al. 2011b). Measurement displaying DIC measurements near the blade root on the low pressure side for a 400 lb. applied static load; (c) blade curvature along section line lengthwise of blade (root shown on left), (d) displacement contour near blade root, (e) strain at points 0 and 1 (shown in b)) as a function of applied load, and (f) overlay of displacement contour with blade image (LeBlanc et al. 2013)

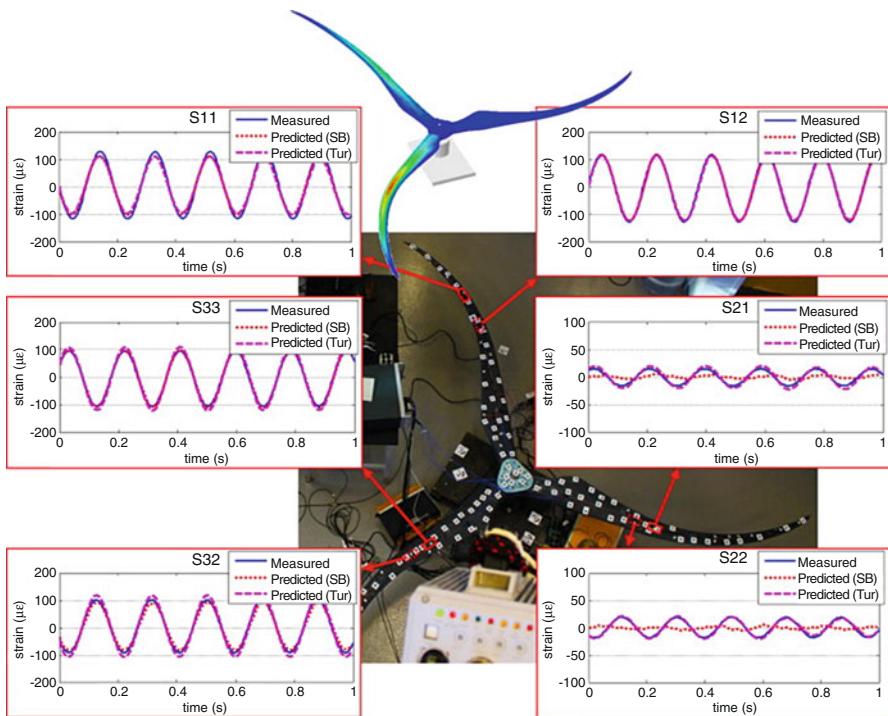
transformation matrices that relate corresponding points in the overlap areas. The rotation and translation matrices are computed by using point cloud registration techniques such as singular value decomposition (SVD) or iterative closest point (ICP) (Salvi et al. 2007). In this research, the authors were able to provide displacement and strain measurements over the entire surface of the blade itself by combining 16 FOVs recorded by a single camera and stitched together with photo-editing software (LeBlanc et al. 2011b).

Other studies performed on the same 9 m blade embedding defects (i.e., wave defects of well-known geometry inserted at specified locations along the blade length) aimed to compare the pros and cons of different sensing techniques (e.g., DIC, shearography, acoustic emission, fiber-optic strain sensing, thermal imaging, and piezoelectric sensing) as SHM tools for detecting the defects and track the resultant damage due to fatigue testing. DIC measurements were able to reveal the areas characterized by higher levels of the strain compared to the surrounding footprint, revealing the location of the defects. During the fatigue tests, the strain amplification in the vicinity of the fault became more apparent confirming that DIC is an effective method for locating cracks and wave defects on a targeted structure (Niezrecki et al. 2014).

In 2012, researchers proposed to use 3DPT to identify the mode shapes of a Southwest Windpower Skystream 4.7 wind turbine blade (Baqersad et al. 2012). In the study, the blade contained several optical targets, and it was excited at different frequencies using a shaker and a pluck test. The main difference with the research performed by Ozbek et al. 2013 was the extraction of mode shapes from time domain data. The blade's response was captured using two high-speed cameras, and an operational modal approach was used to extract mode shapes by transferring the recorded time domain data to the frequency domain. Results of the study demonstrated that DIC can be used as an acquisition system for measurement of low-frequency vibrations.

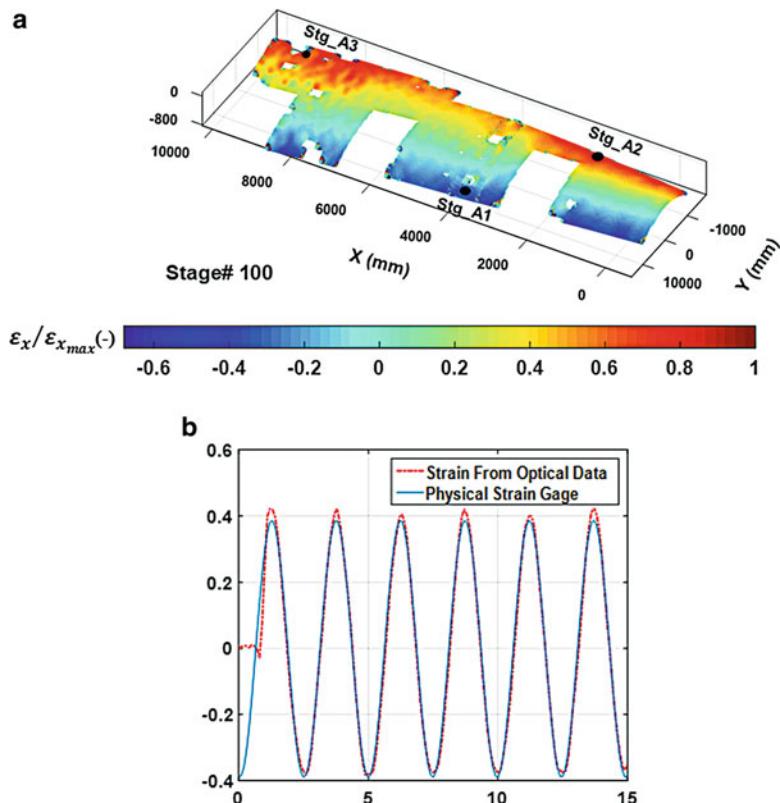
Another set of tests performed on the same Southwest Windpower Skystream blade validated the use of 3D-DIC and 3DPT to predict the dynamic stresses and strains from limited measurement locations using an expansion process in conjunction with finite element models (Carr et al. 2014). In the research, the methods were discussed both during the static and dynamic loadings typical of a wind turbine certification process. In another work, Baqersad et al. studied the use of displacements measured with 3DPT to obtain full-field strain data on the Southwest wind turbine (Baqersad et al. 2015; Baqersad et al. 2016). In their research, measured displacements were transformed to a finite element degree of freedom using a modal expansion algorithm. The expanded displacements were applied to the FEM to calculate the full-field dynamic strain of the structure. The predicted values were compared to measured data by using six mounted strain gages (see Fig. 11).

3DPT was used to measure surface displacements for dynamic events and as a complementary set of data for validating the modal expansion algorithm presented in the research. Between 2014 and 2017, Poozesh et al. started developing a multi-camera system for measuring full-field strain and displacement over fields of view (FOVs) larger than those that can be achieved with a single system. The multi-camera system was used during the quasi-static and fatigue tests that are required during the utility-scale wind turbine blade certification procedure required by the International Electro-Technical Commission standard IEC61400–23 (Poozesh et al. 2017). In this study, the authors first evaluated the potential of 3D-DIC in measuring strain and displacement over a large section of a 50-m utility-scale blade subjected to edgewise quasi-static and cyclic loadings. Then, the authors explored the error associated with using a multi-camera system (i.e., a system composed of two independently calibrated



**Fig. 11** Comparison of the strain measured with mounted strain gages, predicted using modes of the single blade and using the modes of the turbine for a sinusoidal input at the frequency of the first mode of the blade (5.3 Hz) (Baqersad et al. 2015)

stereovision systems) in measuring 3D displacement and extracting structural dynamics parameters on a mock setup emulating a utility-scale wind turbine blade. The idea was validated by performing 3DPT measurements to obtain the displacement fields and operating shapes of a Southwest Windpower Skystream 4.7 wind turbine blade excited using an unknown impact force. Following data collection, the displacement fields were stitched together to reconstruct the global 3D displacement field as a point cloud. The accuracy of the stitching approaches was estimated by comparing the displacement of the overlapped area measured by the two individual systems. Results proved that the time traces for the different targets in the two fields of view were almost identical (i.e., out-of-plane displacement errors equal to  $\sim 200 \mu\text{m}$  with standard deviations less than  $22 \mu\text{m}$ ). Recently, the approach has been further extended to obtain full-field displacement and strain over a  $\sim 10$ -meter-long section (Poozesh et al. 2018) and the modal parameters from an analysis of the displacement of a  $\sim 20$ -meter-long section (Sabato et al. 2018a) of a  $\sim 60$  meter utility-scale wind turbine blade. The obtained results have shown that the proposed system can detect in-plane displacement as low as 2 mm, mechanical strain with an error below 2%, and natural frequencies with an error below 5% when compared with data recorded using traditional wire-based sensors as shown in Fig. 12.



**Fig. 12** Strain results obtained using the multi-camera system for monitoring a ~10-meter-long section of a ~60-meter-long utility-scale wind turbine blade: (a) Full-field strain in the Y direction; (b) comparison of the strain measured with mounted strain gages (blue lines) and the recorded strain using the multi-camera system (red lines) for location Stg\_A1 (Poozesh et al. 2018)

The stitching approach was shown to be useful for two camera pairs; therefore, in principle, it should be able to be extended to three or more sensor pairs, leveraging the use of optical techniques for analyses on large-sized structures.

Besides the use of optically based methods for wind turbine SHM and NDI, DIC has found applications in other energy sectors too. Hohmann et al. proposed this technique to extend the life of civil structures within nuclear power plants. Concrete degradation (i.e., swelling or shape deformation) of containment building is of primary interest for safety reasons. In their study, the researchers validate the use of 3D-DIC to precisely determine the change of shape and strain in a concrete containment vessel at a nuclear power plant during a pressure structural integrity test (Hohmann et al. 2012).

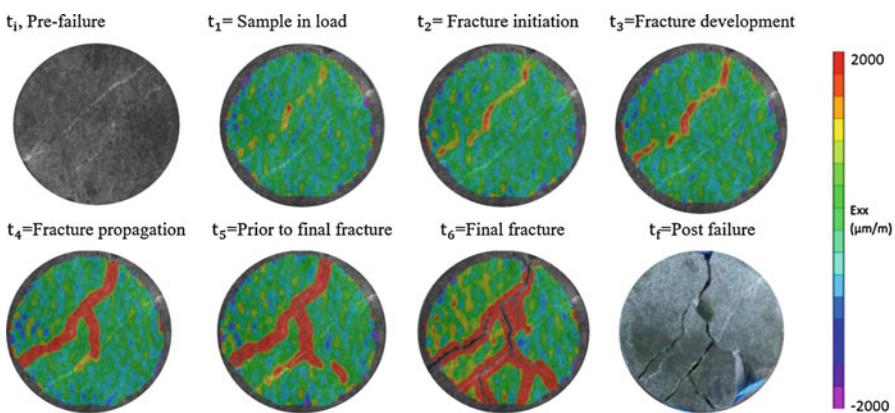
In 2017, researchers at the University of Louisiana at Lafayette started to apply DIC technique to measure strain development and obtain information about the interaction between natural fractures and induced fractures in heterogeneous rock

system (e.g., sandstone sample) for petroleum extraction. The study proved that DIC is capable of illuminating the failure mechanism and strain development in naturally fractured media as shown in Fig. 13 (Mokhtari et al. 2017).

Similar studies were conducted to calculate the strain of elastic rocks and overcome shortfalls of traditional strain measurement. DIC was used to provide information about the strain fields and fracture patterns as different types of sandstones and carbonate formations were undergoing *Brazilian tests*. DIC results were capable of determining winding and erratic fracture initiation and propagation with consistent strain mapping (Nath et al. 2017). These research activities have shown to be promising for petroleum engineers interested in gaining an understanding of the fracturing mechanisms of unusual rock types and their index of brittleness (Parshall et al. 2017).

## NDE and SHM Using DIC in Material Testing Applications

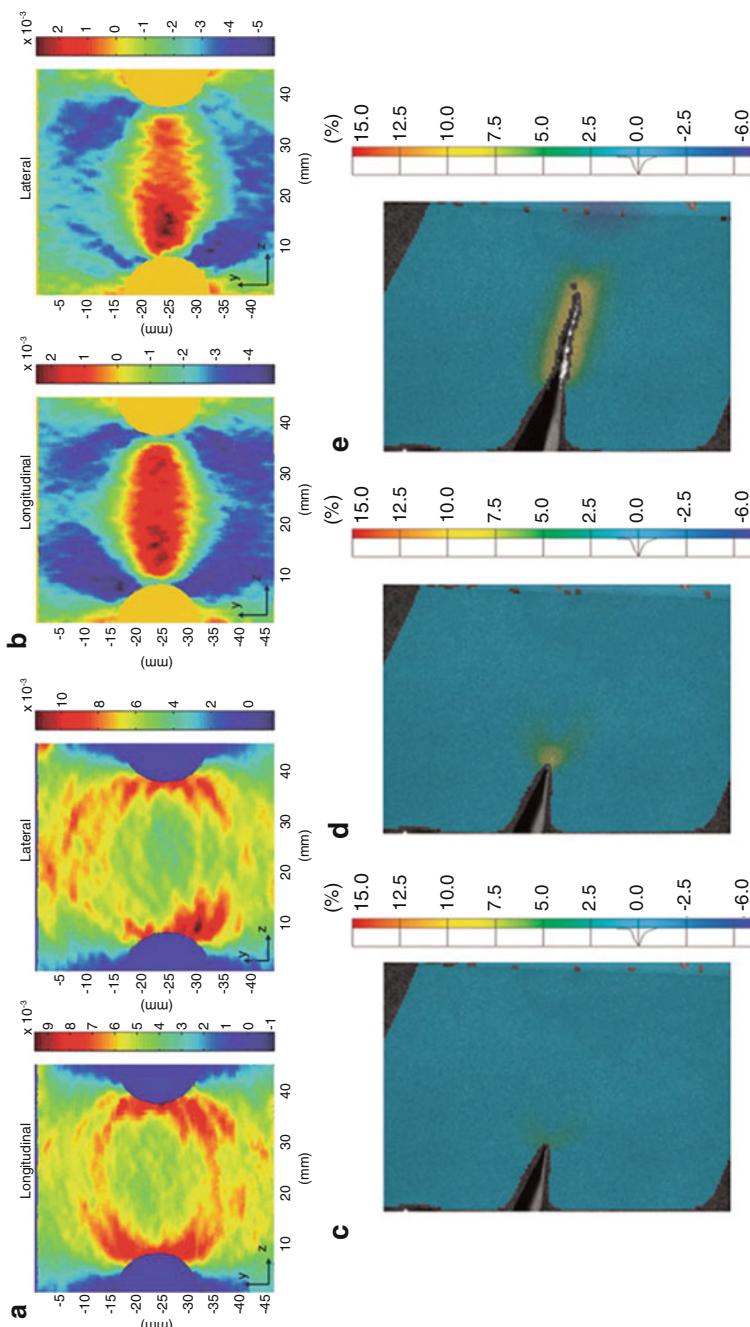
DIC, due to its noncontact nature and the capability to detect displacement at a microscopic scale, has been widely used in testing applications to determine mechanical parameters of diverse materials and experimentally validate the theoretical model of fracture mechanics. In 2002, Abanto-Bueno and Lambros developed a customized DIC code to obtain the in-plane displacement field surrounding a propagating crack, assessing the resistance behavior of functionally graded materials (FGM) and extracting stress intensity factors and resistance curves (Abanto-Bueno and Lambros 2002). The optically based technique was useful for investigating regions of K-dominance (i.e., areas where the theoretical asymptotic fields describe the near-tip deformation) in the materials and displaying the displacement and strain fields generated by the presence of the crack and from possible rigid body motion. The possibility to refer to full-field strain and displacement maps allowed the gaining of better knowledge about the FGM mechanical behavior under cracking conditions.



**Fig. 13** Strain measurement in heterogeneous Buda limestone with natural fractures at different times and fracturing stages (Mokhtari et al. 2017)

In 2005, Lagattu et al. studied the efficiency of DIC for measuring in-plane displacement as high-strain gradients were applied to a composite laminate having a hole in its structure. Results showed that DIC was able to measure increase in strain values up to 12% and demonstrated the efficacy of the method for strain mapping (Lagattu et al. 2004). In 2010, Bernasconi et al. studied the effect of notches and fiber orientation on static and fatigue strength of short glass fiber-reinforced (SGFR) polyamide 6. These materials are of great interest in a perspective of a metal replacement for load-bearing applications (e.g., automotive industry); therefore, gaining an appreciation on how fiber orientation distribution at notch locations affect strain field is fundamental to understand the fatigue strength and failure mechanism of the system. In the study, DIC was applied to two specimens, one injected longitudinally and the other laterally. The goal was to measure the strain fields under increasing controlled tensile loading from zero to 5.4 kN, in 300 N steps, highlighting the different mechanical responses for the two selected injection types as shown in Fig. 14a and b (Bernasconi et al. 2010).

In 2012, researchers from Drexel University performed a comparison of active GUW, passive AE, and DIC to develop a novel SHM approach based on the combination of real-time optical and acoustic NDT to monitor and quantify crack growth in Al 2024 compact tension specimens. By using data fusion techniques that condensed information collected using different NDT methods, the authors defined a damage index. The application of this integrated SHM approach resulted in reliable damage detection and quantified crack growth measurements (Vanniamparambil et al. 2012). The same systems were used by Rouchier et al. in 2013 to monitor fiber-reinforced mortar samples during tensile-loading tests. In the study, DIC was employed to detect microscopic (i.e., as small as 2  $\mu\text{m}$ ) and macroscopic cracks, and their tortuosity, while image processing procedures were conducted to track their evolution in time over the course of the degradation process (Rouchier et al. 2013). Starting in the same year, researchers from the Ecole Centrale de Nantes experimentally monitored fracture process of concrete structures studying crack opening profiles and crack length using DIC (Alam et al. 2012). The analyses of these parameters are the direct indicators of the local failure in the material microstructure and can completely describe the fracture process of the material; therefore, their analyses were of importance for preventing catastrophic failures. A three-point bending test was performed on geometrically similar concrete beams with different sizes and data recorded continuously using two 2-megapixel digital cameras with 75 mm macro lenses located at the two faces of the structure. Following that experience, the same group of researchers started developing data fusion techniques using AE and DIC to identify crack openings and size of a fracture zone in concrete. Analyses of the results showed that AE was used to determine the location of fracture growth due to microcracks and macrocracks; however, DIC was able to quantify the dimension of crack openings at various areas of the crack itself. Also, both techniques combined were able to detect the location of the crack tip, confirming that the fusion of the two methods was useful in identifying the fracture process zone and cracking mechanisms of the targeted system (Alam et al. 2014). Other tests

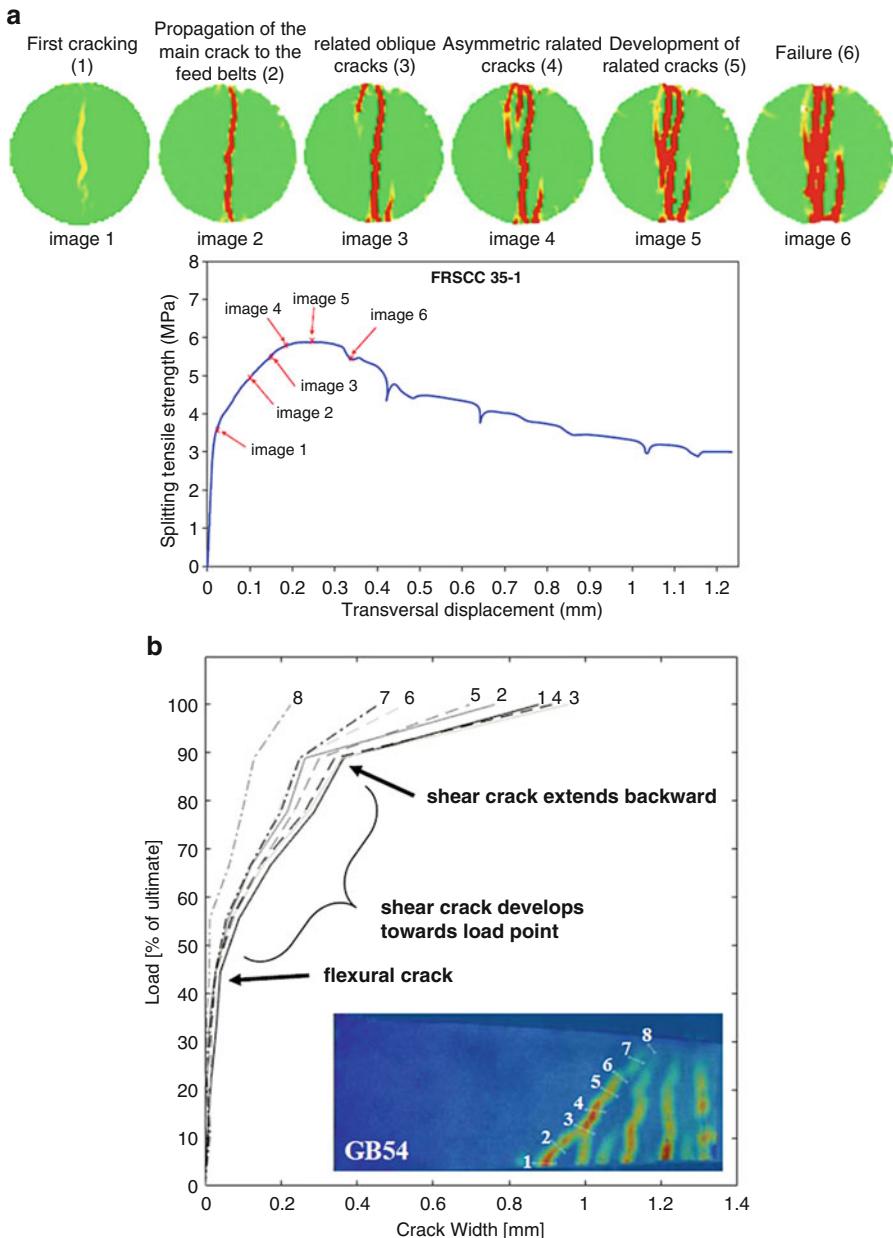


**Fig. 14** Displacement and strain field measured using DIC on a short glass fiber-reinforced (SGFR) polyamide 6 with embedded notches. (a) Axial displacement fields along y direction; (b) transversal strain fields along z direction (Bernasconi et al. 2010), DIC crack growth for Al 2024 specimens during tensile tests: (c) no crack extension, (d) stable, and (e) unstable crack growth (Vanniamparambil et al. 2012)

performed on reinforced concrete specimens, crack openings, and crack spacing experimentally measured by DIC and AE were compared with those determined by simulations using Eurocode2. Results showed that DIC provided a very accurate measurement of surface displacement, defining crack openings and spacing, while AE resulted in information about the internal damages caused by the applied service loading (Alam et al. 2015; Alam and Loukili 2017). Similar studies regarding combined approaches and sensor fusion techniques were also proposed by Annamdas et al. (2014) (steel specimens embedding electrode sparked hemispherical defects on their surfaces using a piezoelectric wafer-based electromechanical impedance technique and a DIC system for multi-crack monitoring) and Omondi et al. (2016) (control of cracks in concrete structures at laboratory standard tests to characterize the mechanical and damage behavior of pre-stressed concrete sleepers using AE and DIC).

Concrete and reinforced concrete were among the most studied materials using optically based techniques. Fayyad and Lees (2014) applied DIC to investigate mode I propagation, crack opening displacement, and crack profiles in reinforced concrete (RC). In the study, the relationship between the fracture properties and the properties of the concrete and steel reinforcement was investigated experimentally performing three-point bending tests on small-scale-reinforced concrete specimens. In 2015, Di Benedetti et al. used 3D-DIC to study the mechanical behavior of fiber-reinforced polymer (FRP) RC beams during a quasi-static four-point bending test and comparison with strain gages and potentiometers. The peculiarity of this study was that two beams patterned with different techniques were used; one using the traditional sprayed speckles, while the other employing computer generated with randomly spaced speckles with a predefined nominal diameter and laser printed on a transfer paper. Analyses of the results showed that 3D-DIC was the only system capable of measuring out-of-plane displacement and vertical displacements measured with DIC agreed well with those recorded by potentiometers. Also, strain values in the compression region of the beam were more accurately estimated when the computer-generated speckle pattern was used, while the sprayed pattern was more accurate in the identification of the neutral axis position (Di Benedetti et al. 2015). In the same year, 3D-DIC was also used to investigate the fracture mechanism of steel FRC concrete in a tensile splitting test as shown in Fig. 15a (Boulekache et al. 2015). Other studies involving the use of 3D-DIC for the purpose of gaining additional insight about the development and degradation of shear resisting mechanisms to calibrate a numerical model were performed by Cholostiakow et al. This research also aimed to examine the size effect and overall shear behavior of GFRP RC beams with and without shear reinforcement. 3D-DIC was employed to monitor the crack openings along the shear crack as shown in the example reported in Fig. 15b (Cholostiakow et al. 2016, 2017).

Considering the increasing use of composite materials such as carbon-fiber-reinforced plastic (CFRP) composites in the aircraft industry, researchers have started to consider scarf-type bonded patches as an alternative mechanical fastened repair technology. Between 2012 and 2013, Caminero et al. used DIC and 3D-DIC to validate the design and certification of repairs carried out in CFRP materials using



**Fig. 15** 3D-DIC for crack monitoring in FRC concrete specimen: (a) fracture mechanism in a tensile splitting test (Boulekbache et al. 2015); (b) shear crack width in a FRP RC beam (Cholostiakow et al. 2016)

scarf patch adhesive bonding as the composite plate was subjected to tensile loading. They determined the risk of debonding of the patch under loading in critical stages of the structure's life. DIC was successfully used to obtain full-field strain measurements in open-hole tensile CFRP specimens, highlighting the presence of localized damages and the capability of highlighting strain fields and high strain concentrations around cracks. 3D-DIC was used for performing damage assessment of a scarf repair under tensile loading and comparison with results obtained from Lamb waves analyses, and FEM showed good agreement between the different datasets (Caminero et al. 2012, 2013a). The same authors performed more investigation on CFRP composites notched (i.e., open-hole) specimens to gain a better understanding of the behavior of mechanical discontinued region that could modify the strength of the material under uniaxial tensile loading. Obtained results confirmed the potential and accuracy of optically based techniques to study possible damage in laminates with discontinuities and adhesively bonded patch repairs. Nevertheless, this technique still leaves unsolved the problem of identifying internal faults (e.g., resin cracking, fiber/matrix interface failure, delamination, and fiber micro-buckling) (Caminero et al. 2013b, 2014).

Starting in 2012, researchers started to focus on the fracture properties of thin aluminum inclusions embedded in anisotropic paperboard composites and performed tensile tests on nonconventional heterogeneous specimens using DIC for recovering information about the metal deformation and the evolution of the damaging processes leading to the detachment of the inclusion from the surrounding laminate composite (Bolzon et al. 2012). The research included the combined comparison of experimental 3D-DIC measurement and analytical FEM techniques focusing on the out-of-plane displacements that develop in tensile tests as fracture propagates across the paperboard composite and during quasi-static inflation experiments (Bolzon et al. 2017a; Zappa et al. 2017). The same research group performed studies to analyze anisotropic damage mechanisms in forged metal alloy specimens, both smooth and notched (Gariboldi et al. 2016), and thin aluminum notched metal foils under uniaxial tensile load (Bolzon et al. 2017b). In the first study, DIC was used to complement the analyses performed using micrograph and was able to gain further information about the strain values on the surface of the notched specimens proving that residual strains were confined to the notch root as well as to the flanges of advanced macrocracks. In the latter, 3D-DIC was used for complementing usual test records, monitoring the displacement distribution, the configuration changes of the specimens, and highlighting wrinkling of the samples during the tests.

In 2017, Asl et al. used 3D-DIC to validate the design of scaled-down wind turbine blades subcomponents made of glass/epoxy materials that emulate the strain field experienced in the full-scale components. 3D-DIC was used to measure the full-field strain distribution and displacement in the bottom flange of the I-beams designed using the similitude theory as the specimen was subjected to a three-point bending test. The measured transverse deflection values were compared to analytical models and FEM simulations (Asl et al. 2017).

## Other Applications and Vision-Based Systems

This section describes the application of vision-based system in areas that are not mentioned in the previous sections.

In 2005, Tong studied the effects of changes in image acquisition conditions (e.g., variable brightness, contrast, uneven local lighting, and blurring) to assess the precision of strain mapping (Tong 2005). In 2009, Hutt and Cawley proposed to use 3D-DIC as a cheap and quick testing technique for the detection of structural defects such as cracking and corrosion. The specific case analyzed in the research focused on a crack at a hole in an aluminum plate. Although strain fields were within the noise floor of the measurement, and consequently not detectable by the system, cracks were successfully detected at moderate loads by measuring the displacement discontinuity across the crack itself. In this test, a minimum displacement of ~0.3 pixels, cracks of lengths higher than 2 mm, and strain field of 180  $\mu\text{m}$  were detected (Hutt and Cawley 2009). In 2016, Malesa et al presented two different strategies (i.e., for overlaying FOVs and for distributed FOVs) for stitching data obtained from multi-camera 3D-DIC systems. They validated the two proposed approaches by performing monitoring of a 12-meter span graded metal plate hall's arch and an analysis of three individual pipelines in a nitrogen plant (Malesa et al. 2016). In 2014, Zappa et al. proposed a novel technique to reduce DIC uncertainty in dynamic tests due to blurriness caused by a translating target. Despite the fact that this research focused on the 2D case, results can be applied to 3D scenarios as well. In their work, the researchers developed two different methods to simulate the motion effect on a reference image. The first one addressed the issue of sub-pixel shifting using the discrete Fourier transform (Reu 2011) to simulate the pure translation of the target, while the second one averages the shifted images to simulate the motion effects. These methods allow the simulation of the acquired images in a real dynamic test and the estimation of the measurement uncertainty caused by the motion effect. Validation performed using harmonic vibration tests showed that good agreement existed between the experiments and the simulations regarding measurement accuracy modification in dynamic conditions. Also, the authors proposed a numerical technique to evaluate the motion effect present in the acquired images. The method used proved to be capable of reducing bias errors and measurement uncertainties enhancing the performances of DIC in dynamic applications (Zappa et al. 2014a, b). The same research group has developed a robust framework for reducing uncertainties of vision-based measurement and evaluated the effects of motion blur. The details of their research are external to the focus of this chapter, but the interested reader can refer to Lavatelli and Zappa (2016, 2017), Zappa and Hasheminejad (2017).

In 2015, Park et al. developed a method to measure 3D structural displacements using a motion capture system that measures the 2D coordinate of a number of markers (e.g., LED lights) with at least three cameras to calculate their 3D coordinates. The effectiveness of the method was tested by comparing the displacements measured in a free vibration experiment of a 3-story structure using both the motion capture system and laser displacement sensors, showing an accuracy of approximately 0.645 mm (Park et al. 2015).

## Future Directions

Optically based measurement techniques have proven to be suitable tools for SHM and NDE of structures, especially as their size increases. Future directions to increase the widespread applicability of DIC may involve investigations aimed to develop novel calibration methodologies to simplify the use of stereovision systems as large-sized structures such as long-deck suspension bridges or utility-scale wind turbine blades need to be monitored or measured. Some researchers have already started to investigate in this direction. For instance, Santos et al. worked on developing a novel in-situ calibration methodology for vision sensors to measure the displacement of long-deck suspension bridges where vertical and transversal movements of the deck were on the order of 1 meter. The minimum requirements for the proposed system consisted of a minimum of two cameras, a set of targets fixed on the structure being monitored, and the knowledge of the distances between those targets to determine the intrinsic and extrinsic parameters needed for the calibration with an accuracy having a standard deviation of two pixels (i.e., in-plane deviation lower than 3.5 mm and out-of-plane deviation smaller than 7.5 mm) (Santos et al. 2012a, b). Based on these findings, the authors proposed an algorithm to perform simultaneously and in real time the vision system calibration and the full motion tracking of large structures. The algorithm showed to be suitable for the estimation of the 6-DOF rotation and translation components of the motion of the structure and the vision system auto-calibration (Santos et al. 2016). Other researchers proposed to determine the extrinsic parameters of the stereovision system by employing sensor board embedding MEMS-based Inertial Measurement Units and Radar systems for calculating the relative position of cameras in space (Sabato et al. 2018b).

Other developments could focus on evolving novel patterning systems. For instance, Mazzoleni et al. proposed a novel toner-transfer technique to impress a well-defined and repeatable speckle pattern on plane and curved surfaces (Mazzoleni et al. 2015), while Park et al. used speckle pattern created by a laser (Park et al. 2018). In 2016, Dworakowski et al. presented a vision-based method based on DIC to measure the deflection curve of a cantilever beam and use those data as an input to a novel detection algorithm based on line segments and voting methods (Dworakowski et al. 2016). Ye et al. developed a vision-based system programmed with three different image processing algorithms (i.e., the grayscale pattern matching (GPM) algorithm, the color pattern matching (CPM) algorithm, and the mean shift tracking (MST) algorithm) for multi-point structural dynamic displacement measurement. The authors validated their approach measuring the displacement influence lines of an arch bridge during a loading test (Ye et al. 2016b). Moreover, interested readers can refer to Ye et al. (2016c) for a review of vision-based systems other than DIC. Other researchers are starting to explore how DIC can be performed using spectrums of light other than the visible spectrum (e.g., infrared). This would enable patterns to be applied to structures without having to have them be visible.

Volumetric DIC is also another area of research that is currently going through significant expansion. The volumetric measurement relies on scans such as MRI or CT scans that obtain images at several layers of depth beyond what is shown at the

surface. Multiple images are taken in layers to quantify the internal deformation and strain of a material.

As camera technology is improving every year and new cameras are offered at lower prices with better resolution and higher frame rates, the vision-based systems are more frequently used for structural health monitoring and nondestructive damage detection. Likewise, the high computational capabilities offered by new computers can help in applying new image processing techniques for improved damage detection.

Many researchers have tried to integrate finite element analysis with digital image correlation to identify damages. As a future direction, the image processing software packages can be integrated into the finite element software programs. This can enable new algorithms for correlation analysis and to use both numerical and experimental techniques to identify damages in structures.

---

## Conclusions

In this chapter, the development of optically based techniques such as digital image correlation for nondestructive evaluation and structural health monitoring has been analyzed and surveyed. For nearly four decades, DIC has been improved by many researchers making it an effective and flexible optical technique for surface deformation measurement, geometry profile detection, full-field displacement and strain analysis, and dynamic measurement, from the macroscopic to microscale. This chapter has focused on the measurement aspects of stereophotogrammetry including DIC and point tracking, which are finding new applications every year. The technology has improved significantly over the years with advances in cameras, computational power, and improved computational algorithms. Due to its noncontact nature, the optical measurement approach can overcome the limitations that characterize traditionally wire-based and contact sensors. Moreover, considering the rapid development of optical sensors and computational algorithms, this technique has become a well-established standard technique for static testing and its use rapidly taking hold for dynamic investigations. The possibility to unify multiple fields of view together and combine the optical measurements with autonomous, unmanned approaches and with finite element models has tremendous potential to provide information and data that could not have been obtained just a few short years ago. These advances and the leveraging of observations of structural features instead of applied patterns will enable DIC to increase its applicability from simple laboratory-scale testing to structural health monitoring and nondestructive evaluation performed on real-world, large-scale engineering structures.

---

## References

- Abanto-Bueno J, Lambros J (2002) Investigation of crack growth in functionally graded materials using digital image correlation. Eng Fract Mech 69(14–16):1695–1711. [https://doi.org/10.1016/S0013-7944\(02\)00058-9](https://doi.org/10.1016/S0013-7944(02)00058-9)
- Alam SY, Loukili A (2017) Transition from energy dissipation to crack openings during continuum-discontinuum fracture of concrete. Int J Fracture 206(1):49–66. <https://doi.org/10.1007/s10704-017-0200-8>

- Alam SY, Loukili A, Grondin F (2012) Monitoring size effect on crack opening in concrete by digital image correlation. *Eur J Environ Civ En* 16(7):818–836. <https://doi.org/10.1080/19648189.2012.672211>
- Alam SY, Saliba J, Loukili A (2014) Fracture examination in concrete through combined digital image correlation and acoustic emission techniques. *Constr Build Mater* 69:232–242. <https://doi.org/10.1016/j.conbuildmat.2014.07.044>
- Alam SY, Loukili A, Grondin F, Rozière E (2015) Use of the digital image correlation and acoustic emission technique to study the effect of structural size on cracking of reinforced concrete, vol 143. *Eng Fract Mech*, pp 17–31. <https://doi.org/10.1016/j.engfracmech.2015.06.03>
- Annamdas VGM, Chew Y, Pang JHL, Hoh HJ, Zhou K, Song B (2014) Fatigue growth analysis of pre induced surface defects using piezoelectric wafer based impedance method and digital image correlation system. *J Nondestruct Eval* 33(3):413–426. <https://doi.org/10.1007/s10921-014-0237-9>
- ASCE (2017) Infrastructure report card. American Society of Civil Engineering. Available online. <https://www.infrastructurereportcard.org/>. Accessed 30 Jan 2018
- Asl ME, Niezrecki C, Sherwood J, Avitabile P (2017) Experimental and theoretical similitude analysis for flexural bending of scaled-down laminated I-beams. *Compos Struct* 176:812–822. <https://doi.org/10.1016/j.compositesct.2017.06.017>
- Baqersad J, Carr J, Lundstrom T, Niezrecki C, Avitabile P, Slattery M (2012) Dynamic characteristics of a wind turbine blade using 3D digital image correlation. In: Proceedings of the SPIE, Health monitoring of structural and biological systems 7979: 79790L, 20 April 2012, San Diego, CA
- Baqersad J, Niezrecki C, Avitabile P (2015) Full-field dynamic strain prediction on a wind turbine using displacements of optical targets measured by stereophotogrammetry. *Mech Syst Signal Process* 62–63:284–295. <https://doi.org/10.1016/j.ymssp.2015.03.021>
- Baqersad J, Poozesh P, Niezrecki C, Avitabile P (2016) A noncontacting approach for full-field strain monitoring of rotating structures. *J Vib Acoust* 138(3):031008. <https://doi.org/10.1115/1.4032721>
- Baqersad J, Poozesh P, Niezrecki C, Avitabile P (2017) Photogrammetry and optical methods in structural dynamics – a review. *Mech Syst Signal Process* 86:17–34 <https://doi.org/10.1016/j.ymssp.2016.02.011>
- Bartilson DT, Wieghaus KT, Hurlebaus S (2015) Target-less computer vision for traffic signal structure vibration studies. *Mech Syst Signal Process* 60:571–582. <https://doi.org/10.1016/j.ymssp.2015.01.005>
- Bernasconi A, Cosmi F, Zappa E (2010) Combined effect of notches and fibre orientation on fatigue behaviour of short fibre reinforced polyamide. *Strain* 46(5):435–445. <https://doi.org/10.1111/j.1475-1305.2009.00667.x>
- Bolzon G, Buljak V, Zappa E (2012) Characterization of fracture properties of thin aluminum inclusions embedded in anisotropic laminate composites. *Frattura ed Integrità Strutturale* (19):20–28. <https://doi.org/10.3221/IGF-ESIS.19.02>
- Bolzon G, Shahmardani M, Liu R, Zappa E (2017a) A combined experimental-numerical investigation of the failure mode of thin metal foils. *Procedia Structural Integrity* 3:168–171. <https://doi.org/10.1016/j.prostr.2017.04.030>
- Bolzon G, Shahmardani M, Liu R, Zappa E (2017b) Failure analysis of thin metal foils. *Frattura ed Integrità Strutturale* 11(42):328–336. <https://doi.org/10.3221/IGF-ESIS.42.34>
- Boulekbache B, Hamrat M, Chemrouk M, Amziane S (2015) Failure mechanism of fibre reinforced concrete under splitting test using digital image correlation. *Mater Struct* 48(8):2713–2726. <https://doi.org/10.1617/s11527-014-0348-x>
- Busca G, Cigada A, Mazzoleni P, Zappa E, Franzini M (2012) Cameras as displacement sensors to get the dynamic motion of a bridge: performance evaluation against traditional approaches. In: Frangopol D (ed) Proceedings of the 6th IABMAS 2012, July 8–12, 2012, Stresa, Italy
- Busca G, Cigada A, Mazzoleni P, Zappa E (2014) Vibration monitoring of multiple bridge points by means of a unique vision-based measuring system. *Exp Mech* 54(2):255–271. <https://doi.org/10.1007/s11340-013-9784-8>
- Calebi M (2000) Seismic instrumentation of buildings. Report No. 00–157, United States Geological Survey

- Caminero MA, Pavlopoulou S, López-Pedrosa M, Nicolaïsson BG, Pinna C, Soutis C (2012) Using digital image correlation techniques for damage detection on adhesively bonded composite repairs. *Adv Compos Lett* 21(2):51–57
- Caminero MA, Pavlopoulou S, Lopez-Pedrosa M, Nicolaïsson BG, Pinna C, Soutis C (2013a) Analysis of adhesively bonded repairs in composites: damage detection and prognosis. *Compos Struct* 95:500–517. <https://doi.org/10.1016/j.compstruct.2012.07.028>
- Caminero MA, Lopez-Pedrosa M, Pinna C, Soutis C (2013b) Damage monitoring and analysis of composite laminates with an open hole and adhesively bonded repairs using digital image correlation. *Compos Part B Eng* 53:76–91. <https://doi.org/10.1016/j.compositesb.2013.04.050>
- Caminero MA, Lopez-Pedrosa M, Pinna C, Soutis C (2014) Damage assessment of composite structures using digital image correlation. *Appl Compos Mater* 21(1):91–106. <https://doi.org/10.1007/s10443-013-9352-5>
- Caprioli A, Manzoni S, Zappa E (2011) People-induced vibrations of civil structures: image-based measurement of crowd motion. *Exp Tech* 35(3):71–79. <https://doi.org/10.1111/j.1747-1567.2009.00574.x>
- Carr J, Baqersad J, Nieszrecki C, Avitabile P (2014) Full-field dynamic strain on wind turbine blade using digital image correlation techniques and limited sets of measured data from photogrammetric targets. *Exp Tech*. <https://doi.org/10.1111/ext.12129>
- Catbas FN, Zaurin R, Gul M, Gokce HB (2011) Sensor networks, computer imaging, and unit influence lines for structural health monitoring: case study for bridge load rating. *J Bridg Eng* 17 (4):662–670
- Chiang CH, Shih MH, Chen W, Yu CP (2011). Displacement measurements of highway bridges using digital image correlation methods. In: Proceedings of the SPIE, Precision Engineering Measurements and Instrumentation 8321: 83211G, 16Nov. 2011, Yunnan, China. <https://doi.org/10.1117/12.904303>
- Cholostiakow S, Di Benedetti M, Guadagnini M, Gowda C, Barros J, Zappa E (2016) Experimental and numerical study on the shear behaviour of geometrically similar FRP RC beams. In: Teng JG, Dai JG (eds) Proceedings of the 8th FRP Composites CICE 2016, The Hong Kong Polytechnic University, pp 1–6
- Cholostiakow S, Di Benedetti M, Guadagnini M, Zappa E (2017) Size effect in FRP RC beams with and without shear reinforcement. In: Procof the FRPRCS 2017, 14–15 October 2017, Anaheim, CA
- Chu TC, Ranson WF, Sutton MA (1985) Applications of digital-image-correlation techniques to experimental mechanics. *Exp Mech* 25(3):232–244
- Chung HC, Liang J, Kushiyama S, Shinozuka M (2004) Digital image processing for non-linear system identification. *Int J Nonlinear Mech* 39(5):691–707. [https://doi.org/10.1016/S0020-7462\(03\)00021-0](https://doi.org/10.1016/S0020-7462(03)00021-0)
- Cigada A, Mazzoleni P, Tarabini M, Zappa E (2013) Static and dynamic monitoring of bridges by means of vision-based measuring system. In: Cunha A (ed) Topics in dynamics of bridges, vol 3. Springer, New York, pp 83–92
- Di Benedetti M, Cholostiakow S, Fergani H, Zappa E, Cigada A, Guadagnini M (2015) 3D-DIC for strain measurement in small scale GFRP RC specimens. In: Proceedings of the SMAR 2015: 1–8, 7–9 September 2015, Antalya, Turkey
- Dizaji MS, Alipour M, Harris DK (2017a) Leveraging vision for structural identification: a digital image correlation based approach. In: Sutton M, Reu PL (eds) International digital imaging correlation society. Springer, New York, pp 121–124. [https://doi.org/10.1007/978-3-319-51439-0\\_29](https://doi.org/10.1007/978-3-319-51439-0_29)
- Dizaji MS, Harris DK, Alipour M, Ozbulut OE 2017b En“vision”ing a novel approach for structural health monitoring – a model for full-field structural identification using 3D-digital image correlation. In: Proceedings of the ISHMII 8, 5–8 December 2017, Brisbane, Australia
- Doebling SW, Farrar CR, Prime MB, Shevitz DW (1996) Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review. Report No. LA-13070-MS, Los Alamos National Laboratory
- Dong CZ, Ye XW, Jin T (2018) Identification of structural dynamic characteristics based on machine vision technology. *Measurement* (in press). <https://doi.org/10.1016/j.measurement.2017.09.043>

- Drewry MA, Georgiou GA (2007) A review of NDT techniques for wind turbines. *Insight-Non-Destructive Test Cond Monit* 49(3):137–141. <https://doi.org/10.1784/insi.2007.49.3.137>
- Drinkwater BW, Wilcox PD (2006) Ultrasonic arrays for non-destructive evaluation: a review. *NDT E Int* 39(7):525–541. <https://doi.org/10.1016/j.ndteint.2006.03.006>
- Du Y, Diaz FA, Burguete RL, Patterson EA (2011) Evaluation using digital image correlation of stress intensity factors in an aerospace panel. *Exp Mec* 51(1):45–57. <https://doi.org/10.1007/s11340-010-9335-5>
- Dworakowski Z, Kohut P, Gallina A, Holak K, Uhl T (2016) Vision-based algorithms for damage detection and localization in structural health monitoring. *Struct Cont Health Monit* 23(1):35–50 <https://doi.org/10.1002/stc.1755>
- Farrar CR, Worden K (2007) An introduction to structural health monitoring. *Philos Trans A Math Phys Eng Sci* 365(1851):303–315
- Fayyad TM, Lees JM (2014) Application of digital image correlation to reinforced concrete fracture. *Proc Mater Sci* 3:1585–1590. <https://doi.org/10.1016/j.mspro.2014.06.256>
- Feng D, Feng MQ (2015) Model updating of railway bridge using in situ dynamic displacement measurement under trainloads. *J Bridg Eng* 20(12):04015019. [https://doi.org/10.1061/\(ASCE\)BE.1943-5592.0000765](https://doi.org/10.1061/(ASCE)BE.1943-5592.0000765)
- Feng D, Feng MQ (2016) Vision-based multipoint displacement measurement for structural health monitoring. *Struc Cont Health Monit* 23(5):876–890. <https://doi.org/10.1002/stc.1819>
- Feng D, Feng MQ (2017) Identification of structural stiffness and excitation forces in time domain using noncontact vision-based displacement measurement. *J Sound Vib* 406:15–28. <https://doi.org/10.1016/j.jsv.2017.06.008>
- Feng D, Feng MQ, Ozer E, Fukuda Y (2015a) A vision-based sensor for noncontact structural displacement measurement. *Sensors* 15(7):16557–16575. <https://doi.org/10.3390/s150716557>
- Feng MQ, Fukuda Y, Feng D, Mizuta M (2015b) Nontarget vision sensor for remote measurement of bridge dynamic response. *J Bridg Eng* 20(12):04015023. [https://doi.org/10.1061/\(ASCE\)BE.1943-5592.0000747](https://doi.org/10.1061/(ASCE)BE.1943-5592.0000747)
- Feng D, Scarangello T, Feng MQ, Ye Q (2017) Cable tension force estimate using novel noncontact vision-based sensor. *Measurement* 99:44–52. <https://doi.org/10.1016/j.measurement.2016.12.020>
- Fu G, Moosa AG (2002) An optical approach to structural displacement measurement and its application. *J Eng Mech* 128(5):511–520. [https://doi.org/10.1061/\(ASCE\)0733-9399\(2002\)128:5\(511\)](https://doi.org/10.1061/(ASCE)0733-9399(2002)128:5(511))
- Gariboldi E, Naumenko K, Ozhoga-Maslovskaja O, Zappa E (2016) Analysis of anisotropic damage in forged Al–cu–mg–Si alloy based on creep tests, micrographs of fractured specimen and digital image correlations. *Mat Sci Eng A* 652:175–185. <https://doi.org/10.1016/j.msea.2015.11.090>
- Gheitasi A, Harris DK, Hansen M (2018) An experimental-computational correlated study for describing the failure characteristics of concrete across two scale levels: mixture and structural component. *Exp Mech* 58(1):11–32. <https://doi.org/10.1007/s11340-017-0319-6>
- Gholizadeh S (2016) A review of non-destructive testing methods of composite materials. *Procedia Structural Integrity* 1:50–57. <https://doi.org/10.1016/j.prostr.2016.02.008>
- Gholizadeh S, Leman Z, Baharudin BTHT (2015) A review of the application of acoustic emission technique in engineering. *SEM* 54(6):1075–1095. <https://doi.org/10.12989/sem.2015.54.6.1075>
- Giurgiutiu V, Cuc A (2005) Embedded non-destructive evaluation for structural health monitoring, damage detection, and failure prevention. *Shock Vib Dig* 37(2):83–105
- Glisic B, Inaudi D (eds) (2008) Fiber optic methods for structural health monitoring. Wiley, Hoboken
- Hanke R, Fuchs T, Uhlmann N (2008) X-ray based methods for non-destructive testing and material characterization. *Nucl Instrum Methods Phys Res A* 591(1):14–18. <https://doi.org/10.1016/j.nima.2008.03.016>
- Helfrick MN, Niezrecki C (2007) An investigation of the use of 3-D optical measurements to perform structural health monitoring. In: Proceedings of the 6th IWSHM, Sept. 11–13, Stanford, CA

- Helfrick MN, Niezrecki C, Avitabile P, Schmidt T (2008) 3D digital image correlation methods for full-field vibration measurement. In: 26th conference and exposition on structural dynamics, vol 1. Curran Associates Inc., Red Hook, pp 87–96
- Helfrick MN, Pingle P, Niezrecki C, Avitabile P (2009a) Using full-field vibration measurement techniques for damage detection. In: 27th conference and exposition on structural dynamics, vol 1. Curran Associates Inc., Red Hook, pp 2239–2249
- Helfrick MN, Pingle P, Niezrecki C, Avitabile P (2009b) Optical non-contacting vibration measurement of rotating turbine blades. In: 27th conference and exposition on structural dynamics, vol 1. Curran Associates Inc., Red Hook, pp 222–226
- Helfrick MN, Niezrecki C, Avitabile P, Schmidt T (2011) 3D digital image correlation methods for full-field vibration measurement. *Mech Syst Signal Process* 25(3):917–927. <https://doi.org/10.1016/j.ymssp.2010.08.013>
- Hohmann BP, Bruck P, Esselman TC, Schmidt T (2012) Digital image correlation (DIC): an advanced nondestructive testing method for life extension of nuclear power plants. *Int Atom IAEA-CN-194*, 14–18 May 2012, Salt Lake City, UT, 43(32).
- Hutt T, Cawley P (2009) Feasibility of digital image correlation for detection of cracks at fastener holes. *NDT&Int* 42(2):141–149. <https://doi.org/10.1016/j.ndteint.2008.10.008>
- Ji YF, Chang CC (2008) Nontarget image-based technique for small cable vibration measurement. *J Bridg Eng* 13(1):34–42 [https://doi.org/10.1061/\(ASCE\)1084-0702\(2008\)13:1\(34\)](https://doi.org/10.1061/(ASCE)1084-0702(2008)13:1(34))
- Jones CA, Reynolds P, Zappa E, Manzoni S, Cigada A (2011) Verification of crowd dynamic excitation estimated from image processing techniques. In: Proulx T (ed) *Dynamics of civil structures*, vol 4. Springer, New York, pp 205–216
- Kahn-Jetter ZL, Chu TC (1990) Three-dimensional displacement measurements using digital image correlation and photogrammic analysis. *Exp Mech* 30(1):10–16
- Khuc T, Catbas FN (2017) Computer vision-based displacement and vibration monitoring without using physical target on structures. *Struct Infrastruc E* 13(4):505–516. <https://doi.org/10.1080/15732479.2016.1164729>
- Kim SW, Kim NS (2013) Dynamic characteristics of suspension bridge hanger cables using digital image processing. *NDT&E Int* 59:25–33 <https://doi.org/10.1016/j.ndteint.2013.05.002>
- Kohut P, Holak K, Uhl T, Ortyl Ł, Owerko T, Kuras P, Kocierz R (2013) Monitoring of a civil structure's state based on noncontact measurements. *SHM* 12(5–6):411–429. <https://doi.org/10.1177/1475921713487397>
- Kohut P, Gaska A, Holak K, Ostrowska K, Śladek J, Uhl T, Dworakowski Z (2014a) A structure's deflection measurement and monitoring system supported by a vision system. *TM-Technisches Messen* 81(12):635–643. <https://doi.org/10.1515/teme-2014-1057>
- Kohut P, Holak K, Uhl T, Maczak J, Szulim P (2014b) Application of vision based damage detection for real civil engineering structure. *Key Eng Mat* 588:22–32 <https://doi.org/10.4028/www.scientific.net/KEM.588.22>
- Kohut P, Holak K, Martowicz A, Uhl T (2017) Experimental assessment of rectification algorithm in vision-based deflection measurement system. *Nondestruct Test Eva* 32(2):200–226. <https://doi.org/10.1080/10589759.2016.1159306>
- Koltsida I, Tomor A, Booth C (2013) The use of digital image correlation technique for monitoring masonry arch bridges. In: Radić J, Kušter M, Šavor Z (eds) *Proceedings of the 7th international conference on ARCH bridges – ARCH 2013*. Zagreb, Croatia
- Lagattu F, Brillaud J, Lafarie-Frenot MC (2004) High strain gradient measurements by using digital image correlation technique. *Mater Charact* 53(1):17–28. <https://doi.org/10.1016/j.matchar.2004.07.009>
- Lavatelli A, Zappa E (2016) Modeling uncertainty for a vision system applied to vibration measurements. *IEEE Trans Instrum Meas* 65(8):1818–1826. <https://doi.org/10.1109/TIM.2016.2541359>
- Lavatelli A, Zappa E (2017) A displacement uncertainty model for 2-D DIC measurement under motion blur conditions. *IEEE T Instrum and Meas* 66(3):451–459. <https://doi.org/10.1109/TIM.2016.2644898>

- LeBlanc B, Niezrecki C, Avitabile P (2010) Structural health monitoring of helicopter hard landing using 3D digital image correlation. In: Proceedings of the SPIE, Health monitoring of structural and biological systems 7650: 76501V, 7–11 March 2010, San Diego, CA
- LeBlanc B, Niezrecki C, Avitabile P, Chen J, Sherwood J, Hughes S (2011a) Full-field inspection of a wind turbine blade using three-dimensional digital image correlation. Proc. SPIE, Industrial and commercial applications of smart structures technologies 8348: 83482I, 29 April 2011, San Diego, CA
- LeBlanc B, Niezrecki C, Avitabile P, Sherwood J, Chen J (2011b) Surface stitching of a wind turbine blade using digital image correlation. In: Allemand R, De Clerck J, Niezrecki C, Blough JR (eds) Topics in modal analysis II, vol 6. Springer, New York, pp 277–284
- LeBlanc B, Niezrecki C, Avitabile P, Chen J, Sherwood J (2013) Damage detection and full surface characterization of a wind turbine blade using three-dimensional digital image correlation. SHM 12(5–6):430–439. <https://doi.org/10.1177/1475921713506766>
- Lee JJ, Shinozuka M (2006) A vision-based system for remote sensing of bridge displacement. Ndt&E Int 39(5):425–431 <https://doi.org/10.1016/j.ndteint.2005.12.003>
- Lundstrom T, Niezrecki C, Avitabile P (2011) Rigid body correction using 3D digital photogrammetry for rotating structures. In: Proulx T (ed) Optical measurements, Modeling, and metrology, vol 5. Springer, New York, pp 307–321. [https://doi.org/10.1007/978-1-4614-0228-2\\_38](https://doi.org/10.1007/978-1-4614-0228-2_38)
- Lundstrom T, Baqersad J, Niezrecki C (2013) Using high-speed stereophotogrammetry to collect operating data on a Robinson R44 helicopter. In: Allemand R, De Clerck J, Niezrecki C, Wicks A (eds) Special topics in structural dynamics, vol 6. Springer, New York, pp 401–410
- Lundstrom T, Niezrecki C, Avitabile P (2015) Appropriate rigid body correction of stereophotogrammetry measurements made on rotating systems. Exp Tech 39(6):25–34. <https://doi.org/10.1111/ext.12030>
- Lundstrom T, Baqersad J, Niezrecki C (2016) Monitoring the dynamics of a helicopter main rotor with high-speed stereophotogrammetry. Exp Tech 40(3):907–919. <https://doi.org/10.1111/ext.12127>
- Luo PF, Chao YJ, Sutton MA, Peters WH (1993) Accurate measurement of three-dimensional deformations in deformable and rigid bodies using computer vision. Exp Mech 33(2):123–132
- Lynch JP (2007) An overview of wireless structural health monitoring for civil structures. Philos Trans A Math Phys Eng Sci 365(1851):345–372
- Malesa M, Szczepanek D, Kujawińska M, Świercz A, Kołakowski P (2010) Monitoring of civil engineering structures using digital image correlation technique. In: Brémand F (ed) Proceedings of the 14th ICEM, Jul. 4–9, Poitiers, France, vol 6, p 31014. <https://doi.org/10.1051/epjconf/20100631014>
- Malesa M, Malowany K, Pawlicki J, Kujawinska M, Skrzypczak P, Piekarzuk A, Lusa T, Zagorski A (2016) Non-destructive testing of industrial structures with the use of multi-camera digital image correlation method. Eng Fail Anal 69:122–134. <https://doi.org/10.1016/j.engfailanal.2016.02.002>
- Mazzoleni P, Zappa E (2012) Vision-based estimation of vertical dynamic loading induced by jumping and bobbing crowds on civil structures. Mech Syst Signal Pr 33:1–12. <https://doi.org/10.1016/j.ymssp.2012.06.009>
- Mazzoleni P, Zappa E, Matta F, Sutton MA (2015) Thermo-mechanical toner transfer for high-quality digital image correlation speckle patterns. Opt Laser Eng 75:72–80. <https://doi.org/10.1016/j.optlaseng.2015.06.009>
- McCormick N, Lord J (2012) Digital image correlation for structural measurements. Proc Inst Civ Eng 165(4):185
- McGinnis MJ, Pessiki S, Turker H (2005) Application of three-dimensional digital image correlation to the core-drilling method. Exp Mech 45(4):359. <https://doi.org/10.1007/BF02428166>
- McGinnis MJ, Smith B, Holloman M, Lisk M, O'Donnell A, Kurama YC (2012) 3-D digital image correlation – an underused asset for structural testing. In: Carrato J, Burns J (eds) Structures congress. ASCE, Chicago, pp 1958–1969. <https://doi.org/10.1061/9780784412367.172>

- Mokhtari M, Hayatdavoudi A, Nizamutdinov R, Rizvi H, Nath F (2017) Characterization of complex fracture propagation in naturally fractured formations using digital image correlation technique. Proc. SPE hydraulic fracturing technology conference and exhibition, 24–26 January, The Woodlands, TX
- Murray CA, Take WA, Hoult NA (2014) Measurement of vertical and longitudinal rail displacements using digital image correlation. *Can Geotech J* 52(2):141–155. <https://doi.org/10.1139/cgj-2013-0403>
- NASA (2009) Chopper drop tests new technology. <https://www.nasa.gov/topics/aeronautics/features/helio-droptest.html>. Accessed 9 Apr 2018
- NASA (2017) NASA crashes helicopter to test safety improvements. <https://www.nasa.gov/larc/nasa-crashes-helicopter-to-test-safety-improvements>. Accessed 9 Apr 2018
- Nath F, Salvati PE, Mokhtari M, Seibi A, Hayatdavoudi A (2017) Observation of fracture growth in laminated sandstone and carbonate rock samples under Brazilian testing conditions using digital image correlation technique. In: Proceedings of the SPE hydraulic fracturing technology conference and exhibition, 4–6 October, Lexington, KY
- Nieuzeck C, Avitabile P, Warren C, Pingle P, Helfrick M (2010) A review of digital image correlation applied to structure dynamics. *AIP Conference Proceedings* 1253(1):219–232 <https://doi.org/10.1063/1.3455461>
- Nieuzeck C, Avitabile P, Chen J, Sherwood J, Lundstrom T, LeBlanc B, Hughes S, Desmond M, Beattie A, Rumsey M, Klute SM, Pedrazzani R, Werlink NJ (2014) Inspection and monitoring of wind turbine blade-embedded wave defects during fatigue testing. *SHM* 13(6):629–643. <https://doi.org/10.1177/1475921714532995>
- Nonis C, Nieuzeck C, Yu TY, Ahmed S, Su CF, Schmidt T (2013) Structural health monitoring of bridges using digital image correlation. In: Proceedings of the SPIE, Health monitoring of structural and biological systems 8695: 869507, 17 April 2013, San Diego, CA. <https://doi.org/10.1117/12.2009647>
- Olaszek P (1999) Investigation of the dynamic characteristic of bridge structures using a computer vision method. *Measurement* 25(3):227–236. [https://doi.org/10.1016/S0263-2241\(99\)00006-8](https://doi.org/10.1016/S0263-2241(99)00006-8)
- Olson L, Barrows D, Abrego A, Burner A (2010) Blade deflection measurements of a full-scale UH-60A rotor system. 2010 AHS specialists' conference on aeromechanics, San Francisco, CA, January 20–22, 2010
- Omondi B, Aggelis DG, Sol H, Sitters C (2016) Improved crack monitoring in structural concrete by combined acoustic emission and digital image correlation techniques. *SHM* 15(3):359–378. <https://doi.org/10.1177/1475921716636806>
- Ozbek M, Rixen DJ (2013) Operational modal analysis of a 2.5 MW wind turbine using optical measurement techniques and strain gauges. *Wind Energy* 16(3):367–381. <https://doi.org/10.1002/we.1493>
- Ozbek M, Meng F, Rixen DJ, ToorenMJL V (2011) Identification of the dynamics of large wind turbines by using photogrammetry. In: Proulx T (ed) Structural dynamics and renewable energy, vol 1. Springer, New York, pp 351–359
- Ozbek M, Meng F, Rixen DJ (2013) Challenges in testing and monitoring the in-operation vibration characteristics of wind turbines. *Mech Syst Signal Process* 41(1–2):649–666. <https://doi.org/10.1016/j.ymssp.2013.07.023>
- Pan B, Tian L, Song X (2016) Real-time, non-contact and targetless measurement of vertical deflection of bridges using off-axis digital image correlation. *Ndt&EInt* 79:73–80. <https://doi.org/10.1016/j.ndteint.2015.12.006>
- Park SW, Park HS, Kim JH, Adeli H (2015) 3D displacement measurement model for health monitoring of structures using a motion capture system. *Measurement* 59:352–362. <https://doi.org/10.1016/j.measurement.2014.09.063>
- Park K, Torbol M, Kim S (2018) Vision-based natural frequency identification using laser speckle imaging and parallel computing. *Comput Aided CivInf* 33(1):51–63. <https://doi.org/10.1111/mice.12312>

- Parshall J, Whitfield S, Jacobs T (2017) Digital image correlation: a new way to look at hydraulic fracturing. *JPT* 69(5):4–5. <https://doi.org/10.2118/0517-0022-JPT>
- Paulsen US, Erne O, Moeller T, Sanow G, Schmidt T (2009) Wind turbine operational and emergency stop measurements using point tracking videogrammetry. In: SEM annual conference and exposition, vol 1. Curran Associates Inc, Red Hook, pp 1128–1138
- Peddle J, Goudreau A, Carlson E, Santini-Bell E (2011) Bridge displacement measurement through digital image correlation. *Bridge Struct* 7(4):165–173. <https://doi.org/10.3233/BRS-2011-031>
- Peters WH, Ranson WF (1982) Digital imaging techniques in experimental stress analysis. *Opt Eng* 21(3):213427–213427. <https://doi.org/10.1117/12.7972925>
- Peters WH, Ranson WF, Sutton MA, Chu TC, Anderson J (1983) Application of digital correlation methods to rigid body mechanics. *Opt Eng* 22(6):226738–226738. <https://doi.org/10.1117/12.7973231>
- Pieraccini M (2013) Monitoring of civil infrastructures by interferometric radar: a review. *Sci World J.* <https://doi.org/10.1155/2013/786961>
- Poozesh P, Baqersad J, Niezrecki C, Avitabile P, Harvey E, Yarala R (2017) Large-area photogrammetry based testing of wind turbine blades. *Mech Syst Signal Process* 86:98–115. <https://doi.org/10.1016/j.ymssp.2016.07.021>
- Poozesh P, Sabato A, Sarrafi A, Niezrecki C, Avitabile P (2018) A multiple stereo-vision approach using three dimensional digital image correlation for utility-scale wind turbine blades, Proc. IMAC XXXVI, 12 February 2018, Orlando, FL
- Reagan D, Sabato A, Niezrecki C, Yu TY, Wilson R (2016) An autonomous unmanned aerial vehicle sensing system for structural health monitoring of bridges. In: Proceedings of the SPIE, Nondestructive characterization and monitoring of advanced materials, aerospace, and civil infrastructure, 9804: 980414. 22 April 2016, Las Vegas, NV. <https://doi.org/10.1117/12.2218370>
- Reagan D, Sabato A, Niezrecki C (2017a) Unmanned aerial vehicle acquisition of three-dimensional digital image correlation measurements for structural health monitoring of bridges. In: Proceedings of the SPIE, Nondestructive characterization and monitoring of advanced materials, aerospace, and civil infrastructure, 10169: 016909, 19 April 2017, Portland, OR. <https://doi.org/10.1117/12.2259985>
- Reagan D, Sabato A, Niezrecki C (2017b) Feasibility of using digital image correlation for unmanned aerial vehicle structural health monitoring of bridges. SHM (in press). <https://doi.org/10.1177/1475921717735326>
- Rehman SKU, Ibrahim Z, Memon SA, Jameel M (2016) Nondestructive test methods for concrete bridges: a review. *Constr Build Mater* 107:58–86. <https://doi.org/10.1016/j.conbuildmat.2015.12.011>
- Reu PL (2011) Experimental and numerical methods for exact subpixel shifting. *Exp Mech* 51:443–452 <https://doi.org/10.1007/s11340-010-9417-4>
- Ribeiro D, Calçada R, Ferreira J, Martins T (2014) Non-contact measurement of the dynamic displacement of railway bridges using an advanced video-based system. *EngStruct* 75:164–180. <https://doi.org/10.1016/j.engstruct.2014.04.051>
- Rizzo P (2014) Sensing solutions for assessing and monitoring railroad tracks. In: Wang ML, Lynch JP, Sohn H (eds) *Sensor technologies for civil infrastructures*, vol 2. Woodhead Publishing, Cambridge, pp 497–524
- Rouchier S, Foray G, Godin N, Woloszyn M, Roux JJ (2013) Damage monitoring in fibre reinforced mortar by combined digital image correlation and acoustic emission. *Constr Build Mater* 38:371–380. <https://doi.org/10.1016/j.conbuildmat.2012.07.106>
- Sabato A, Niezrecki C (2017a) Feasibility of digital image correlation for railroad tie inspection and ballast support assessment. *Measurement* 103:93–105 <https://doi.org/10.1016/j.measurement.2017.02.024>
- Sabato A, Niezrecki C (2017b) Full-scale damage detection of railroad crossties using digital image correlation. In: Sutton M, Reu PL (eds) *International digital imaging correlation society*. Springer, New York, pp 117–120. [https://doi.org/10.1007/978-3-319-51439-0\\_28](https://doi.org/10.1007/978-3-319-51439-0_28)

- Sabato A, Niezrecki C, Fortino G (2017a) Wireless MEMS-based accelerometer sensor boards for structural vibration monitoring: a review. *IEEE Sensors J* 17(2):226–235. <https://doi.org/10.1109/JSEN.2016.2630008>
- Sabato A, Beale CH, Niezrecki C (2017b). A novel optical investigation technique for railroad track inspection and assessment. In: Proceedings of the SPIE, Nondestructive characterization and monitoring of advanced materials, aerospace, and civil infrastructure, 10169: 101692C. <https://doi.org/10.1117/12.2257831>
- Sabato A, Poozesh P, Avitabile P, Niezrecki C (2018a) Experimental modal analysis of a utility-scale wind turbine blade using a multi-camera approach. In: Proceedings of the 13th AIVELA, June 19–22, Ancona, Italy
- Sabato A, Reddy N, Khan S, Niezrecki C (2018b) A novel camera localization system for extending three-dimensional Digital Image Correlation measurement. In: Proceedings of the SPIE, Non-destructive characterization and monitoring of advanced materials, aerospace, and civil infrastructure, 1 April 2018, Denver, CO
- Salvi J, Matabosch C, Fofi D, Forest J (2007) A review of recent range image registration methods with accuracy evaluation. *Image Vis Comput* 25(5):578–596. <https://doi.org/10.1016/j.imavis.2006.05.012>
- Santos CA, Costa CO, Batista JP (2012a) Calibration methodology of a vision system for measuring the displacements of long-deck suspension bridges. *Struc Cont Health Monit* 19(3):385–404. <https://doi.org/10.1002/stc.438>
- Santos CA, Costa CO, Batista JP (2012b) Long deck suspension bridge monitoring: the vision system calibration problem. *Strain* 48(2):108–123. <https://doi.org/10.1111/j.1475-1305.2011.00803.x>
- Santos CA, Costa CO, Batista JP (2016) A vision-based system for measuring the displacements of large structures: simultaneous adaptive calibration and full motion estimation. *Mech Syst Signal Pr* 72:678–694. <https://doi.org/10.1016/j.ymssp.2015.10.033>
- Schmidt T, Tyson J, Galanulis K (2003a) Pull-field dynamic displacement and strain measurement using advanced 3D image correlation photogrammetry: part I. *Exp Tech* 27(3):47–50. <https://doi.org/10.1111/j.1747-1567.2003.tb00115.x>
- Schmidt T, Tyson J, Galanulis K (2003b) Full-field dynamic displacement and strain measurement using advanced 3D image correlation photogrammetry: part II. *Exp Tech* 27(4):22–26. <https://doi.org/10.1111/j.1747-1567.2003.tb00118.x>
- Schneider O (2005) Analysis of SPR measurements from HART II. *Aerosp Sci Technol* 9:409–420. <https://doi.org/10.1016/j.ast.2005.01.013>
- Seo J, Hu JW, Lee J (2015) Summary review of structural health monitoring applications for highway bridges. *J Perform Constr Fac* 30(4):04015072. [https://doi.org/10.1061/\(ASCE\)CF.1943-5509.0000824](https://doi.org/10.1061/(ASCE)CF.1943-5509.0000824)
- Shull PJ (ed) (2016) Nondestructive evaluation: theory, techniques, and applications. CRC Press, New York
- Sirohi J, Lawson MS (2012) Measurement of helicopter rotor blade deformation using digital image correlation. *Opt Eng* 51(4):043603. <https://doi.org/10.1117/1.OE.51.4.043603>
- Sladek J, Ostrowska K, Kohut P, Holak K, Gąska A, Uhl T (2013) Development of a vision based deflection measurement system and its accuracy assessment. *Measurement* 46(3):1237–1249. <https://doi.org/10.1016/j.measurement.2012.10.021>
- Sohn H, Farrar CR, Hemez FM, Czarnecki JJ (2002) A review of structural health review of structural health monitoring literature 1996–2001. Report No. LA-UR-02-2095, Los Alamos National Laboratory
- Stasicki B, Boden F (2009), Application of high-speed videography for in-flight deformation measurements of aircraft propellers. In Proc. 28th International Congress on High-Speed Imaging and Photonics, 9–14 November 2008, Canberra, Australia
- Sutton MA, Wolters WJ, Peters WH, Ranson WF, McNeill SR (1983) Determination of displacements using an improved digital correlation method. *Image Vis Comput* 1(3):133–139 [https://doi.org/10.1016/0262-8856\(83\)90064-1](https://doi.org/10.1016/0262-8856(83)90064-1)

- Sutton MA, Orteu JJ, Schreier H (eds) (2009) Image correlation for shape, motion and deformation measurements: basic concepts, theory and applications. Springer, New York
- Tchakoua P, Wamkeue R, Ouhrouche M, Slaoui-Hasnaoui F, Tameghe TA, Ekemb G (2014) Wind turbine condition monitoring: state-of-the-art review, new trends, and future challenges. *Energies* 7(4):2595–2630. <https://doi.org/10.3390/en7042595>
- Tong W (2005) An evaluation of digital image correlation criteria for strain mapping applications. *Strain* 41(4):167–175. <https://doi.org/10.1111/j.1475-1305.2005.00227.x>
- Udod VA, Van Y, Osipov SP, Chakhlov SV, Usachev EY, Lebedev MB, Temnik AK (2016) State-of-the art and development prospects of digital radiography systems for nondestructive testing, evaluation, and inspection of objects: a review. *Russ J Nondestruct* 52 (9):492–503
- Usamentiaga R, Venegas P, Guerediaga J, Vega L, Molleda J, Bulnes FG (2014) Infrared thermography for temperature measurement and non-destructive testing. *Sensors* 14(7):12305–12348. <https://doi.org/10.3390/s140712305>
- Vanniamparambil PA, Bartoli I, Hazeli K, Cuadra J, Schwartz E, Saralaya R, Kontsos A (2012) An integrated structural health monitoring approach for crack growth monitoring. *J Intel Mat Syst Str* 23(14):1563–1573. <https://doi.org/10.1177/1045389X12447987>
- Vanniamparambil PA, Khan F, Hazeli K, Cuadra J, Schwartz E, Kontsos A, Bartoli I (2013) Novel optico-acoustic nondestructive testing for wire break detection in cables. *Struc Cont Health Monit* 20(11):1339–1350. <https://doi.org/10.1002/stc.1539>
- Warren C, Niezrecki C, Avitabile P (2011a) Determination of wind turbine operating deflection shapes using full-field 3D point-tracking. In: Proulx T (ed) Rotating machinery, structural health monitoring, shock and vibration, vol 5. Springer, New York, pp 217–226
- Warren C, Niezrecki C, Avitabile P (2011b) FRF measurements and mode shapes determined using image-based 3D point-tracking. In: Proulx T (ed) Modal analysis topics, vol 3. Springer, New York, pp 243–252
- Warren C, Niezrecki C, Avitabile P, Pingle P (2011c) Comparison of FRF measurements and mode shapes determined using optically image based, laser, and accelerometer measurements. *Mech Syst Signal Process* 25(6):2191–2202. <https://doi.org/10.1016/j.ymssp.2011.01.018>
- Wu LJ, Casciati F, Casciati S (2014) Dynamic testing of a laboratory model via vision-based sensing. *Eng Struct* 60:113–125. <https://doi.org/10.1016/j.engstruct.2013.12.002>
- Ye XW, Ni YQ, Wai TT, Wong KY, Zhang XM, Xu F (2013) A vision-based system for dynamic displacement measurement of long-span bridges: algorithm and verification. *Smart Struct Syst* 12(3–4):363–379. [https://doi.org/10.12989/ssb.2013.12.3\\_4.363](https://doi.org/10.12989/ssb.2013.12.3_4.363)
- Ye XW, Su YH, Han JP (2014) Structural health monitoring of civil infrastructure using optical fiber sensing technology: a comprehensive review. *Sci World J.* <https://doi.org/10.1155/2014/652329>
- Ye XW, Dong CZ, Liu T (2016a) Force monitoring of steel cables using vision-based sensing technology: methodology and experimental verification. *Smart Struct Syst* 18(3):585–599. <https://doi.org/10.12989/ssb.2016.18.3.585>
- Ye XW, Dong CZ, Liu T (2016b) Image-based structural dynamic displacement measurement using different multi-object tracking algorithms. *Smart Struct Syst* 17(6):935–956. <https://doi.org/10.12989/ssb.2016.17.6.935>
- Ye XW, Dong CZ, Liu T (2016c) A review of machine vision-based structural health monitoring: methodologies and applications. *J Sensors* 2016. <https://doi.org/10.1155/2016/710303>
- Yoneyama S, Kitagawa A, Kitamura K, Kikuta H (2005) Deflection distribution measurement of steel structure using digital image correlation. *Proceedings of the SPIE, Optical Diagnostics* 5880: 5880G, 18 August 2015, San Diego, CA
- Yoneyama S, Kitagawa A, Iwata S, Tani K, Kikuta H (2007) Bridge deflection measurement using digital image correlation. *Exp Techniques* 31(1):34–40. <https://doi.org/10.1111/j.1747-1567.2006.00132.x>
- Zappa E, Hasheminejad N (2017) Digital image correlation technique in dynamic applications on deformable targets. *Exp Tech* 41(4):377–387. <https://doi.org/10.1007/s40799-017-0184-3>

- Zappa E, Mazzoleni P, Matinmanesh A (2014a) Uncertainty assessment of digital image correlation method in dynamic applications. *Opt Laser Eng* 59:140–151. <https://doi.org/10.1016/j.optlaseng.2013.12.016>
- Zappa E, Matinmanesh A, Mazzoleni P (2014b) Evaluation and improvement of digital image correlation uncertainty in dynamic conditions. *Opt Laser Eng* 59:82–92. <https://doi.org/10.1016/j.optlaseng.2014.03.007>
- Zappa E, Liu R, Bolzon G, Shahmardani M (2017) High resolution non-contact measurement techniques for three-dimensional deformation processes of paperboard laminates. *Mater Tod Proc* 4(5):5872–5876. <https://doi.org/10.1016/j.matpr.2017.06.061>
- Zaurin R, Catbas FN (2009) Integration of computer imaging and sensor data for structural health monitoring of bridges. *Smart Mater Struct* 19(1):015019. <https://doi.org/10.1088/0964-1726/19/1/015019>
- Zaurin R, Catbas FN (2011) Structural health monitoring using video stream, influence lines, and statistical analysis. *SHM* 10(3):309–332. <https://doi.org/10.1177/1475921710373290>
- Zhang Z (2000) A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell* 22(11):1330–1334. <https://doi.org/10.1109/34.888718>
- Zhou HF, Dou HY, Qin LZ, Chen Y, Ni YQ, Ko JM (2014) A review of full-scale structural testing of wind turbine blades. *Renew Sust Energ Rev* 33:177–187. <https://doi.org/10.1016/j.rser.2014.01.087>



# Structural Health Monitoring

44

Bianca Weihnacht, Uwe Lieske, Tobias Gaul, and Kilian Tschöke

## Contents

Introduction .....	1592
Pipes and Vessels .....	1593
Guided Waves .....	1593
Acoustic Emission Testing .....	1594
Electrical Resistance .....	1596
Composites .....	1597
Carbon Fiber-Reinforced Plastic (CFRP) .....	1597
Glass Fiber-Reinforced Plastic (GFRP) .....	1599
Offshore Structures .....	1600
Underwater Welded Seams .....	1600
Grouted Joints .....	1602
Sensor Networks for SHM Applications .....	1603
Wireless Technologies and Standards .....	1604
Examples of Wireless Sensors for SHM and Condition Monitoring Applications .....	1606
Summary .....	1607
Cross-References .....	1607
References .....	1607

### Abstract

SHM techniques are well suited to permanently monitor structures in order to prevent sudden failure by fatigue behavior and to reduce risks that are connected with that. The sensors are usually permanently installed on the structures. Furthermore, it is necessary to install a reliable data acquisition and data transfer system and to address especially the long-term stability and reliability. This chapter focuses on several application areas. First, the focus lies on pipes and

B. Weihnacht (✉) · U. Lieske · T. Gaul · K. Tschöke

Fraunhofer Institute for Ceramic Technologies and Systems IKTS, Dresden, Germany

e-mail: [bianca.weihnacht@ikts.fraunhofer.de](mailto:bianca.weihnacht@ikts.fraunhofer.de); [uwe.lieske@ikts.fraunhofer.de](mailto:uwe.lieske@ikts.fraunhofer.de);

[tobias.gaul@ikts.fraunhofer.de](mailto:tobias.gaul@ikts.fraunhofer.de); [kilian.tschoeke@ikts.fraunhofer.de](mailto:kilian.tschoeke@ikts.fraunhofer.de)

vessels with damages such as corrosion and cracks. Next, composites are hard to inspect from the outside regarding internal delaminations and defects. SHM measurements ensure alarms in case the integrity of the structure is not assured. Large offshore structures like foundations are hard to access during long periods of time during the year. Therefore, a monitoring system would give permanent information about defects occurring during operation. Last but not least, wireless communication algorithms are described to make installation easy and cost inefficient.

## Introduction

Structural Health Monitoring (SHM) is a continuous or periodic automated method for the determination and monitoring of the state of the object under surveillance within the condition monitoring (according to ISO 17359 in Germany). This is carried out by measurements of permanently installed, e.g., integrated sensors, and by the analysis of the measurement data. Therefore, the techniques differ from classic NDT measurements. First, the sensors and actuators have to be permanently mounted and have to provide reliable results over long periods of time. Reliability is therefore a key issue for SHM. Furthermore, the sensors and actuators stay permanently at the structure and cannot be reused. This leads to commercial obligations which often lead to smaller number of transducers, cheaper production and assembly technologies, etc. Another key issue is therefore the sophisticated data processing for reliable information about the structure, which has to be applied to data of underdetermined sensor networks with very poor signal to noise ratios. It also has to be stated that in general the resolution of SHM methods is therefore lower than the resolution in NDT measurements. If irregularities are found, NDT measurements are a possible next step for a clear classification and localization.

The first steps for the implementation of a monitoring system are the setting up of a concept supported by simulations, the mounting of the components, the measurements, and the data acquisition and management as well as the permanent access to the data and the data transfer. The data has then to be processed to identify damages by indications which are extracted from the measurement data. Often simulations can provide valuable extra information. The goal is the prognosis of the development and the determination of the residual lifetime. The last step is the decision support system to give the operator of the structure valid information for necessary actions.

In general, there are four steps as guideline for SHM measurements as shown in Fig. 1.



**Fig. 1** Principle of a SHM measurement

For the concept phase, information about geometry including hot spots, expected monitoring periods, and other available data are included into a monitoring concept which includes all necessary information such as sensor positions and date communication. This might be supported by additional simulations for the optimal sensor positions. In this phase, a close cooperation between all involved parties is necessary to achieve the best possible solution.

During the monitoring phase after installation, data are collected and transferred for processing. This might be realized online or offline. Data collection needs appropriate data management, redundant data storage, and basic data processing to evaluate the quality. Usually these demands lead to a necessary permanent data access via remote access.

In order to identify damages, the data has to be processed to get information about irregularities and damages and their severity and location. This is commonly achieved either by direct evaluation of the data (e.g., strain measurements) and comparison with acceptable data from standards or by extended imaging algorithms (e.g., for tomography data).

These data will be evaluated during the decision phase preferably by, e.g., structural mechanical engineers who can compare the data to design limits and provide the operator of the structure with recommendations regarding any further operation and/or necessary repairs. This phase also includes prognosis algorithms and risk assessment.

---

## Pipes and Vessels

The use of Structural Health Monitoring techniques for the inspection of pipes and vessels is important due to the large extension of structures and limited access because of insulation and structural features. Nowadays there exist different monitoring systems with different operating principles. Commonly used techniques are those based on active generated guided waves, acoustic emission, and changes in the electrical resistance. All three methods have their specific characteristics and are suitable for different tasks. The following section presents the different methods with their specific characteristics.

### Guided Waves

Guided waves (GW) are often considered for pipes and vessels monitoring due to their ability to propagate several meters in steel with low attenuation. Guided waves possess multimodal behavior, which means that at a given frequency multiple wave modes with different directions of oscillation can propagate at the same time. Guided waves are also dispersive, and their propagation speed is a function of frequency and wall thickness. For further information see ► [Chap 5, “Guided Wave Testing”](#).

Commonly, antisymmetric and shear wave modes are used for the detection of cracks, pittings and corrosion. Because of their dominant particle motion, they show

the best interaction with such kind of damage types as expected. Guided waves oscillate over the whole cross section of a structure (in contrast to surface waves) and are capable of detecting volume discontinuous. Well-known techniques for guided waves in pipes and vessels are sensor rings with EMAT (electromagnetic acoustic transducer) (García-Gómez et al. 2018) and piezoelectric transducers (Lovstad and Cawley 2011). Piezoelectric transducer rings contain between approx. 12 to 24 sensors in multiple rows and are used for excitation of shear waves in a frequency range below 100 kHz. The excited shear wave mode is nondispersive and therefore has less attenuation in comparison to other wave modes. In pipelines this mode travels more than 30 m and is capable of detecting cracks and corrosion areas due to the reflected signal amplitude. Another application for piezoelectric-based transducers is generating wave modes with high out-of-plane particle motions (commonly called antisymmetric wave modes). Therefore the A0 Lamb wave is the primary wave mode used. It can be used for local monitoring of welded seams. In comparison to the shear wave mode, for the relevant frequency range. Permanently installed monitoring systems underlie changes in environmental conditions, like temperature, moisture, or pressure. To prevent false alarms, these variations have to be compensated (Croxford et al. 2010).

Another widely used application is the determination of integral wall thickness between two sensors. Plane piezoelectric transducers can also be used for the excitation of symmetric Lamb wave modes. In a certain frequency range, the symmetric wave mode shows a strong dispersive behavior, and its propagation velocity is highly dependent on frequency and on wall thickness. With a constant frequency, the wave velocity of the symmetric wave mode depends only on the wall thickness, and the arrival time of the symmetric wave package gives the average wall thickness of the traveled area.

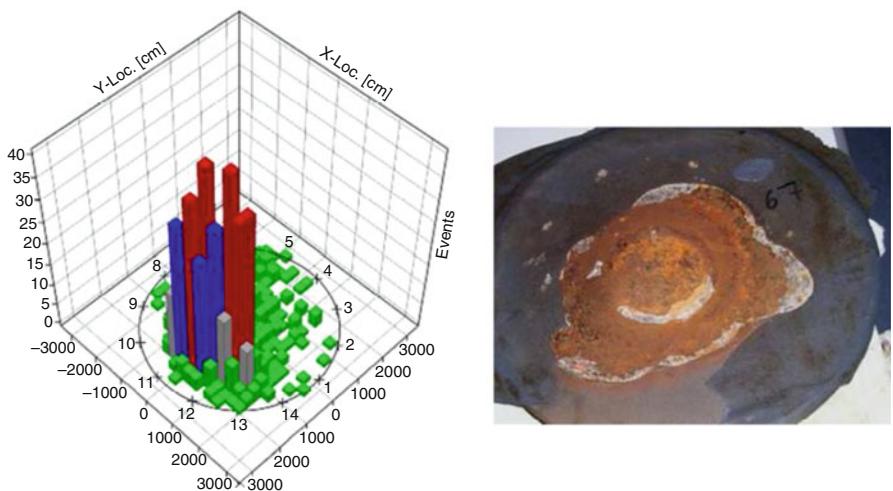
## Acoustic Emission Testing

Another method for monitoring extensive structures based on guided waves is the acoustic emissions (AE) technique. The differences to the previous section technique arise from the mechanism of wave generation. While guided waves are generated actively by a transducer with a specific frequency and shape on the structure, an acoustic emission is generated through the rapid relaxation of localized stress energy inside the structure. Typical sources of acoustic emission are friction noise, crack growth, and plastic deformation due to aging, temperature gradients, or external mechanical forces. Depending on the formation mechanism of the acoustic emission, the frequency range is typically between 1 kHz and 1 MHz. For more detailed information on acoustic emission, see ▶ [Chap. 4, “Acoustic Emission.”](#)

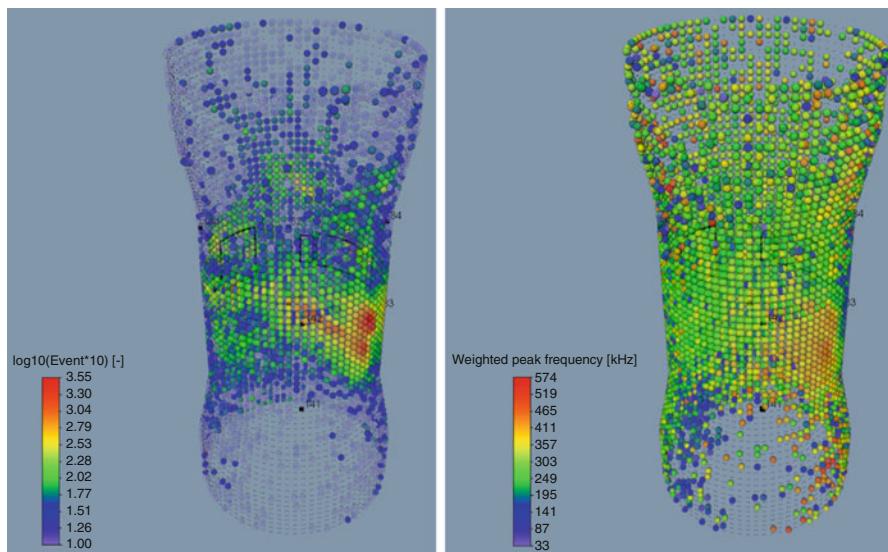
A well-known application of acoustic emission is the monitoring of flat tank bottoms for corrosion damage and leakage. The testing is performed without pressurization of the tank, and the acoustic waves propagate through the liquid product to the sensors mounted on the outer tank wall. With a frequency-domain-based pattern recognition, the measured acoustic emissions are grouped into the two desired

sources, corrosion and leakage damage. An additional source location of acoustic events shows the specific position of detected corrosion and leakage damage, as shown in Fig. 2. In contrast to acoustic waves in metallic plates, which disperse during propagation, acoustic waves propagating through liquid do not disperse. Hence, the application of pattern recognition to measuring data is nearly free of influence by the propagation of the acoustic waves.

Besides corrosion and leakage detection, the acoustic emission method is also capable of detecting the beginning and growth of cracks. Measurements during creep tests at multiple pipe sections show the applicability of acoustic emission at temperatures up to 600 °C. A network of eight distributed sensors was used to monitor multiple pipe sections. Location-dependent plots of different AE parameters show the progress of damage and allow the characterization of the damage mechanism. Representative AE parameters are shown in Fig. 3, where the frequency distribution of events shows areas with high acoustic activity and the weighted peak frequency allows the distinction of source mechanism. A special sensor housing for high temperatures uses a Peltier element to cool the sensor and the sensor electronics. With welded waveguides the permanent coupling between surface and sensor were ensured. Therefore, no contact between the hot surface and the transducer exists. Otherwise, the piezoelectric element would depolarize as the Curie temperature would be exceeded. Currently, available standard piezo-ceramics can stand up to 350 °C as peak temperature and 175 °C at long-term usage without depolarization. In addition, different material combinations exist which raise these temperature limits at the expense of efficiency (Zhang and Fu 2011).



**Fig. 2** Indications from acoustic emission testing and related source (wear plate) (Tscheliesnig et al. 2016)



**Fig. 3** Results of acoustic emission testing after 1 year of dynamic testing at 600 °C. Located events are color-coded in logarithmic scale (left) and the weighted peak frequency during heating period (right)

## Electrical Resistance

Corrosion detection is one of the most challenging tasks in operating pipes and vessels. Environmental conditions and flowing or stored liquids and gases result in a constant loss of material due to chemical or electrochemical processes. Common corrosion rate sensors use the electrical resistance (ER) technique, which is one of the most widely used methods to measure metal loss due to corrosion (Bell and Moore 2007). It has been in use in industrial applications for over 40 years and is a proven online measurement technique in wet and dry conditions. An exposed ferric element in corrosive environment will experience metal loss due to corrosion and consequently see a change (increase) in its electrical resistance. The ER method compares this change to a sealed reference element. Probe elements are made of a material similar to that of the pipe or vessel in which it is placed, in order to simulate as closely as possible the corrosive environment. By taking periodic readings over a fixed time interval, the rate of metal loss can be determined, and a corrosion rate can be calculated. The probe is typically placed in close proximity to the exposed element of interest so that this element is subjected to exactly the same environmental conditions as the reference element (metal resistance is affected by temperature). It is not practical to use an entire pipe as the exposed element. Consequently, a coupon from the pipe of interest (or a coupon of the same type of material) is used. For ER measurements a wide variety of probes with different characteristics for different applications exist. The life span of all probes is inversely proportional to their sensitivity. A probe becomes more sensitive as the element thickness is

reduced, and therefore the time interval in which it can measure a change in resistance decreases, providing faster response. However, in a highly corrosive environment, a thinner element will degrade faster, thereby reducing the life of the probe. A clear understanding of the environment being monitored is needed and will help in selecting a probe that will optimize the life span/response-time decision. Commercially available probes usually withstand temperatures up to 500 °C and external pressures up to 600 bars. The achievable resolution of a probe is up to 0.1% of its span.

---

## Composites

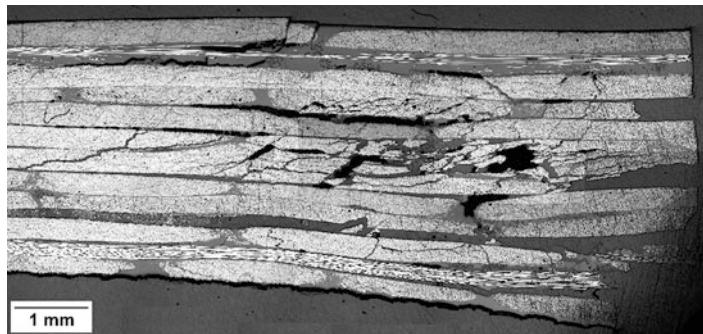
The term “composites” in this chapter refers to fiber-reinforced plastics. These are composite materials made of a polymer matrix reinforced with fibers, usually carbon fibers or glass fibers. The following two sections present two advanced NDE methods (Structural Health Monitoring techniques) based on ultrasonic guided waves to evaluate the condition of composite elements. These methods are known as the acousto-ultrasonic (AU) technique and the acoustic emission (AE) technique. The AU method is presented for carbon fiber-reinforced plastics (CFRP) and the AE method for glass fiber-reinforced plastics (GFRP). Nevertheless, both techniques are applicable to both kinds of composites. Moreover, both methods address the application of condition monitoring during operation. A technique to evaluate the condition of CFRP during fabrication or for quality control is e.g., the high frequency eddy current technique, cf. (Heuer et al. 2015) and ► Chaps. 20, “High-Frequency Eddy Current Techniques” or ► 19, “Eddy Current Testing.”

### Carbon Fiber-Reinforced Plastic (CFRP)

Carbon fiber-reinforced plastics (CFRP) are increasingly used in different applications, such as aerospace, automotive engineering and naval engineering. They are very attractive due to their high stiffness and strength to weight ratios and their corrosion resistance. One of their disadvantages is that composite laminates suffer from delaminations and matrix and fiber crack after impact damages, the so-called barely visible impact damages (BVID); see Fig. 4.

Impact damages are usually investigated and evaluated by conventional ultrasonic impulse-echo techniques which are well established for the NDE of CFRP-components. These techniques are on the one hand time-consuming, and on the other hand, qualified service personnel is required to interpret the signal responses based on expert knowledge and experience. Another disadvantage of impulse-echo techniques is that direct access to the surface of the structure is required. This cannot be ensured for some components, e.g., inside an airplane, without removing all internal cladding.

In structures with a finite cross section, e.g., plates or pipes, so-called guided waves are able to propagate; see ► Chap 5, “Guided Wave Testing”. These guided



**Fig. 4** Barely visible damage after impact (Schubert et al. 2009)

waves travel as Lamb waves and as shear horizontal (SH) waves. Lamb waves are vertically polarized, whereas SH waves are horizontally polarized, see also (Giurgiutiu 2008) for a more detailed description. At any given values of the frequency of the ultrasonic wave signal and the thickness of the plate a multitude of Lamb wave modes may exist. Different guided wave modes are able to propagate in the plate and to interact with several types of damages. Guided waves interact with delaminations and other impact damages such as matrix or fiber cracks.

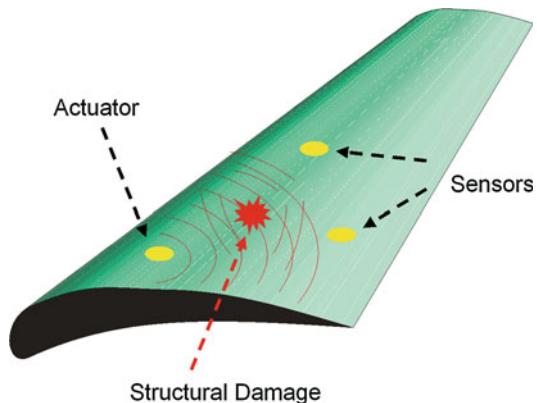
And propagate over a large distance, and they interact over the whole cross section of a specimen. Thus, the surveillance of areas with limited accessibility is possible. Especially the fundamental S0 and A0 modes are of great interest for technical AU applications because of their propagation capability at relatively low-frequency-thickness products.

Figure 5 shows a schematic drawing of an AU measurement system consisting of a network of piezoelectric transducers. It is possible to apply the transducers either at the surface of the monitored object using adhesives or to integrate the transducers within the composite. Due to degradation phenomena of the adhesives, it is recommended to integrate the transducers. Further information can be found in, e.g., Roellig et al. (2012).

All transducers can serve as actuators or sensors. The actuator emits propagating waves within the composite structure. These Lamb waves interact with existing damages and can be recorded at all transducer positions. Signal evaluation is possible either in echo mode (backscattered waves are evaluated at the sending transducer) or in a pitch-catch measurement by several other transducers and the use of a historical baseline. For both possibilities it is important to use the piezoelectric transducers in nonresonant mode.

The interaction with delaminations is characterized as follows. The in-plane component of the SH0 mode shows no interaction. The strongest interaction can be observed by the out-of-plane components of the S0 and A0 modes. Mode conversion (S0 to A0) can be observed during the traveling of the S0 mode through an impact damage. If the wave speed of the propagating guided wave mode is known, it is possible to localize a structural damage using the time shift between the received signals at several transducers. The size of the damage can be determined

**Fig. 5** Schematic drawing of the acousto-ultrasonic technique



by quantitative parameters of the received amplitude or phase change in Lamb wave propagation caused by interaction with the damage. The mentioned parameters correlate with the size of the damaged area. For low impact energies an almost linear decrease in the correlation coefficient can be observed, cf. (Buethe et al. 2016).

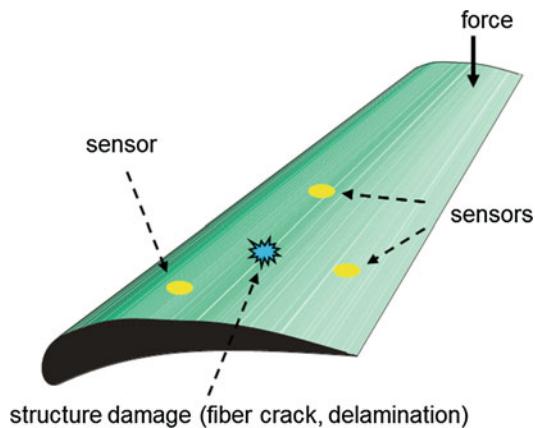
### Glass Fiber-Reinforced Plastic (GFRP)

Glass fiber-reinforced plastics (GFRP) are widely used in different applications such as house building, storage tanks, automotive engineering, and wind turbine rotor blades. The last-mentioned example is chosen to illustrate the acoustic emission (AE) monitoring technique. For more information about acoustic emission, e.g., for CFRP used in automotive engineering, cf. (Schubert et al. 2010; Hoenig et al. 2016), see ▶ Chap. 4, “Acoustic Emission.”

Monitoring of wind turbine rotor blades is usually based on measurements of eigenfrequencies of the blade and the quantifying of changes over time (a.k.a. operational modal analysis) (Pankoke et al. 2011). These systems are mainly focused on ice detection but also on damage detection. Another common monitoring technique for GFRP is strain measurement using resistance strain gauges or even polymer optical fibers, cf. (Bachmann et al. 2008), and see ▶ Chap. 18, “Optical Fiber Methods in Nondestructive Evaluation.” The alternative use of acoustic techniques like AU and AE are also helpful for rotor blade monitoring. Details about AU can be found in the section “Carbon Fiber-Reinforced Plastic”; this section is focused on AE.

Figure 6 shows a transducer network used for AE measurements. A number of piezoelectric transducers are distributed across the rotor blade to form a sensor network for damage monitoring and damage location. The transducers are located in areas of the blade where failure often occurs. The transducers work as a structure-borne sound sensor. Acoustic signals caused by failures in the blade are constantly recorded and evaluated regarding criteria like energy content and travel time. In case of e.g. lightning issue in rotor blades or emi in industrial environment, optical fibre solutions are well suited for both data communication and for the power supply, cf. (Frankenstein et al. 2012).

**Fig. 6** Schematic drawing of the acoustic emission technique



One disadvantage of AE technique is the large number of transducers that are necessary for localizing defects. Because of the high damping of GFRP laminates, the maximum distance between transducers is limited to 3–5 m depending on the laminate thickness. Moreover, the main challenges when applying AE are the strong ambient noise level and the unfavorable acoustic properties of the composites which lead to a separation of desired and disturbed signals. Due to the high damping (up to 10 dB/m) of the GFRP, a transducer mesh can only detect low-frequency Lamb waves, particularly the A0 component. A high sensitivity of the transducers in this frequency domain is required. The localization accuracy is limited because of the high dispersion of the asymmetric wave modes and the velocity inhomogeneities due to varying material thickness. Thus, an error diagnosis has to be integrated within the source localization algorithm.

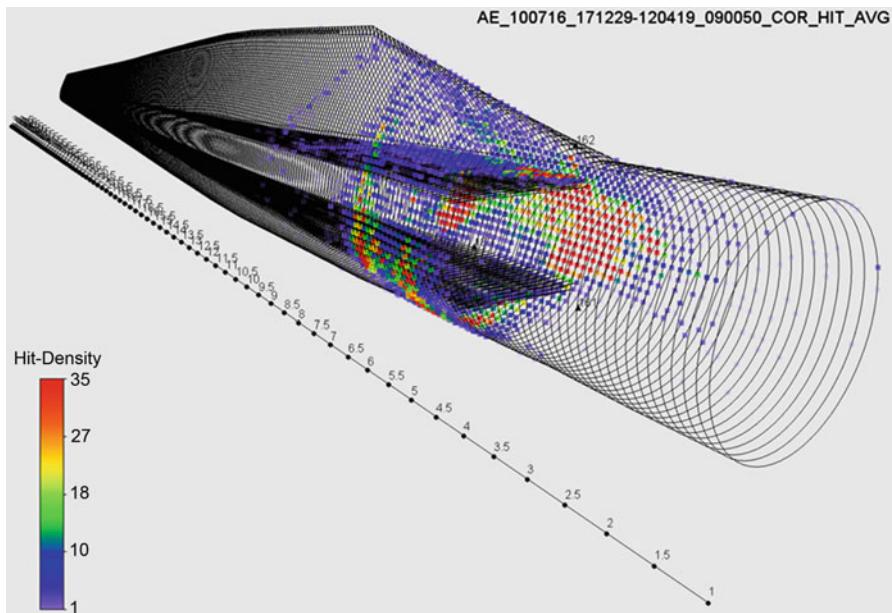
Commonly applied algorithms for AE are based on the arrival times and the choice of the channels around the trigger. The acoustic events are identified by the burst duration amplitude. From characteristic values such as duration and energy content, it is possible to draw conclusions on the type of individual events. A drawing of a 3D-model of the observed specimen is possible by calculating the source points of the acoustic emissions in plane coordinates and a retransformation of these coordinates into a 3D-model. Results of such a reconstruction are shown in Fig. 7 for a rotor blade.

Acoustic emission events generated by a crack are detectable due to the increasing fatigue and the loss of adhesion between the web and the aerodynamic shell (debonding). Acoustic emissions strongly correlate with fatigue of and damages in GFRP.

## Offshore Structures

### Underwater Welded Seams

Offshore structures are regularly inspected regarding their integrity, stability, and safety. Corrosion is of great importance in the offshore sector for all offshore structures due to the harsh environment conditions. Beside cracks initiated



**Fig. 7** Acoustic emission detection plot on a GFRP rotor blade

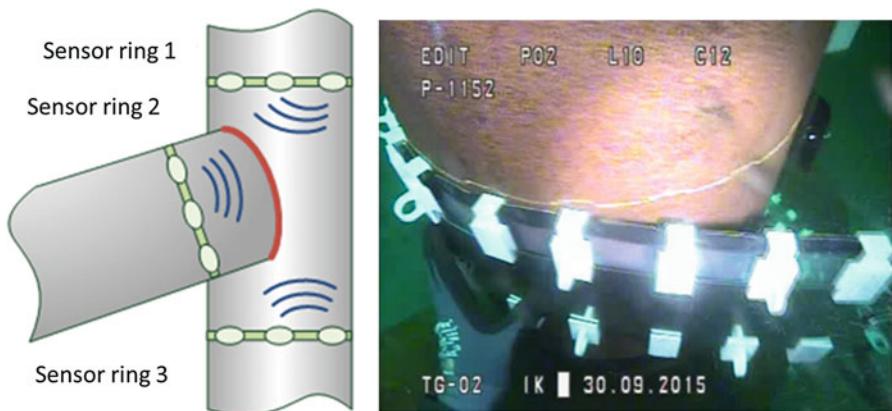
by corrosion mechanisms, surface corrosion is also an issue. Therefore, active corrosion detection sensors are applied on the structure to minimize the corrosion by applying electrochemical negative metals with sufficient activity to work as cathode. Alternatively, this can also be implemented by a direct current from an external power source.

For welded seams ultrasonic techniques and eddy current techniques have been established in the past to determine defect location and size. Furthermore, guided waves and acoustic emission are also appropriate for this task.

For crack detection and the permanent monitoring of crack growth on complete circumferential welds, ultrasonic sensor rings based on guided waves are well suited (Fig. 8), see also ► Chap 5, “Guided Wave Testing”. For this technique, permanently mounted sensors and actuators are located on investigated seams and can be applied in sea and fresh water. Before measuring, a calibration has to be carried out.

Another technique for crack detection of underwater welded seams on monopiles is the acoustic emission technique. A sensor chain with vertical sensor distances of about 3–4 m (depending on the geometry) is placed in the middle of the monopile (Fig. 9).

The monitoring offers a statement regarding the intensity and the activity of irregularities like cracks in welded seams and their over-time and local behavior. Using this technique, it is typically not possible to gain explicit information on the size and the position of the defect (global approach).



**Fig. 8** Sensor ring for welded seam monitoring. One or more rings are applied in areas of high loads. Actuators generate the signal which is detected by sensors after passing the area of interest. From the received data, the defect location and its size is evaluated by tomographic data processing

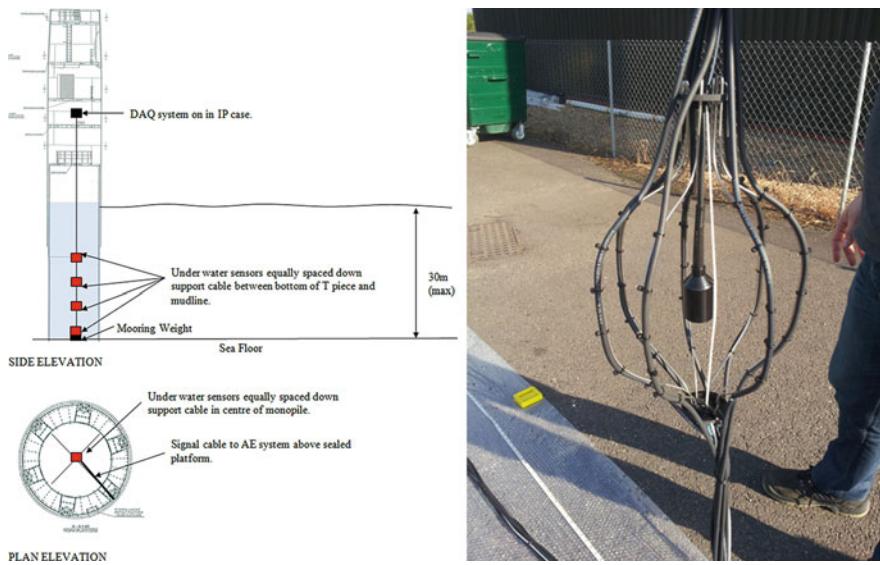
## Grouted Joints

Grouted joints are currently monitored by periodic visual inspection from the topside if accessible. If the grout is situated underwater, continuous monitoring systems can be well suited for the task. This refers to techniques like continuous frequency and tilt measurements. The results give information about changes in frequency and tilt angle and may warn before failure at an early stage.

Another possibility of monitoring above sea level is the measurement of the relative movement of transition pieces and monopiles by using inductive or strain-based sensors. If limits are extended, e.g., by ship collision, the grout has to be evaluated. Grouted joints with bending loads like monopiles are recommended to be monitored every  $120^\circ$  (three sensor systems for the entire pile), whereas grouted joints with mainly tensile and pressure loads (e.g., tripods) are to be monitored at least at one location on the circumference of the grouted joint.

If monitoring systems have given an alert, ultrasonic techniques like guided waves can be applied. These methods are well suited for the detection of delaminations of the grout (interface steel – concrete) and for crack detection within the grout. In order to do so, actuators and sensors are applied, and ultrasonic waves are emitted from the actuators to the sensors. From the reflected wave at the discontinuity, which is detected by sensors located around the circumference of the structural component, information about the condition of the grout can be extracted by tomographic algorithms.

Acoustic emission is also applied for bridges. The applied sensors are sensitive regarding high frequency vibrations in a frequency range of 20–300 kHz. The sensors are sensitive to the development and the growing of defects like cracks, spallings, and fraction due to cyclic loads at the boundary of the defect. The sensors



**Fig. 9** Tool for welded seam inspection on offshore monopiles from the inside (MISTRAS Group)

are only sensitive to larger volumes and therefore give only global information. Single hits can be located if simultaneously detected by several sensors.

## Sensor Networks for SHM Applications

As pointed out in the previous sections, SHM can be utilized for different types of damage detection on buildings, bridges, industrial installations or in aerospace applications to prevent collapse or allow simplified or scheduled maintenance. Many of these applications are typically addressed by wired SHM systems, where distributed sensors are connected to a data interrogator by some kind of a field bus or just plain analog signals are transferred by wire. However, having a large industrial installation might limit the possibilities to implement wired sensors or sensor networks in this way, due to installation and operational efforts.

Wireless sensor networks (WSN) are embedded in the Internet of Things (IoT) context to allow remote surveillance of infrastructure or to automate maintenance inspections.

The following section will give a brief overview of existing wireless technologies and standards. After that a few commercial examples of SHM or condition monitoring solutions will be presented. More detailed information on wireless communications and WSN can be found in the literature (e.g., Gratton 2013; Hersistent et al. 2012; Goldsmith 2005).

## Wireless Technologies and Standards

While planning installation of wireless sensors or wireless sensor networks for SHM applications, the following questions have to be answered:

1. What is the measurement period?
2. Can the measured data be processed on the sensor node or is transfer of raw data required to evaluate a measurement result?
3. Is accurate time synchronization between different sensor nodes during measurement in a network required?

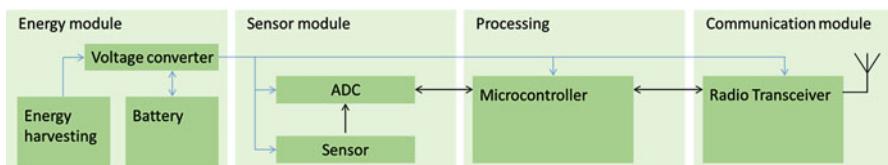
The answers to these questions will define the requirements for a wireless sensor node or sensor network solution.

Wireless sensor nodes are typically build up as illustrated in the following schematic (Fig. 10).

All these components can be combined in one single package or can be used as different building blocks if some modularity of the system is desired. The energy module provides the power for all other system components and can be completely self-sufficient by battery and/or energy harvesting. The sensor module will be designed by the requirements of the measurement task, e.g., if high precision ADC is required like in AET applications or just a temperature or a voltage has to be logged with less accuracy. The same can be applied for the choice of the microcontroller. Highly sophisticated processing of the measurement data will require more powerful microcontrollers. The choice of the communication module will have the most significant impact on the time synchronization and the data transfer within the sensor network.

The usage of radio frequencies is part of national and international regulations. For broad applicability, the usage of free frequency bands with common international regulations like the Industrial, Scientific, and Medical (ISM) frequency bands is common for most companies in the field of wireless sensor networks. As radio bands are a shared medium, interference between different transmitters can happen. Therefore multiplexing methods have to be applied (Table 1).

Beside proprietary technologies, there are several standards available that are suitable for the buildup of wireless sensor networks. The following table gives an overview of the most relevant standards/technologies and their main properties (Table 2).



**Fig. 10** Schematic of wireless sensor nodes (according to Kuorilehto et al. 2007)

**Table 1** Common multiplexing methods

Time division multiplexing (TDM)	Frequency division multiplexing (FDM)	Code division multiplexing (CMD)
Transmitters can access the communication channel in statically or dynamically assigned time slots	The available frequency band is divided in several channels; transmitters can communicate simultaneously on different channels	The data stream is coded to allow reconstruction of the data stream at the receiver. When different codes are used, multiple communications can be handled simultaneously on the same frequency band

**Table 2** Standards and technologies

Standard/technology	Description
IEEE 802.11	This standard defines a set of protocols that are commonly known as Wi-Fi (WLAN). The primary usage is in single-hop environments with high data rates. As low-power operation wasn't a design criterion, the relatively high power consumption limits its usage for wireless sensor networks
Bluetooth	This standard is used for single-hop communication in the 2,4 GHz frequency band. The latest revision of this standard from 2009 (Bluetooth Low Energy) has optimized the power consumption. As the network requires a permanently active master node, the usage for scalable, wireless sensor networks is limited
IEEE 802.15.4	This standard defines two physical layers in the 868 MHz (915 MHz) and the 2,4 GHz frequency band with data rates of 20kbit/s and 250kbit/s, respectively. The network requires a master node. The 802.15.4 MAC Protocol has shown some reliability issues with varying data traffic
ZigBee	This standard was developed for monitoring and control applications and allows star, mesh and cluster tree networks, self-organization, and subsequent routing. An active router node has to be always in range; therefore a permanently available energy source is recommended
UWB	Communication technology that uses a broad frequency spectrum for high data rates for distances up to 70 m with low-power potential and high immunity over narrow band interference. Limited availability of transceivers on the market
TSCH	As IEEE 802.15.4 couldn't allow highly scalable sensor networks for industrial applications, an extension with multiple alternative MAC layers was specified. Advantages are high robustness and reliability with high data rates using multiple channels and consequent time slotted channel hopping (TSCH)
WirelessHART, ISA100.11a	WirelessHART is a synchronized, robust, and scalable network stack. It's an extension to the wired HART field bus to allow wireless components to be implemented in existing industrial installations. The networks stack allows mesh networks. ISA100.11a additionally allows cluster tree topologies

If the application requires different sensor nodes to operate simultaneously, for example, for acoustic emission testing, additional aspects of the sensor nodes have to be considered. Time synchronization is the attempt to address the following problems:

1. Compensation of inaccuracy of crystal clocks of the individual sensor nodes
2. Compensation and elimination of delays during exchange of the actual time within the sensor network

The first aspect can be addressed by the usage of precise crystal clocks and adequate compensation algorithms that consider temperature and aging of the crystal. The second aspect has to be addressed by the time correction procedure within the sensor networks that can be handled by external time references (e.g., GPS) or by distributing the actual time within the network by precise calculation of the time delays by one-way or two-way (NTP, PTP) synchronization algorithms.

## **Examples of Wireless Sensors for SHM and Condition Monitoring Applications**

Several solutions of wireless sensor networks (WSN) are available for home and industrial use. Beside commercial sensor networks solutions for specific applications, some operating systems for wireless sensor nodes have been developed to allow easier access to sensor networks.

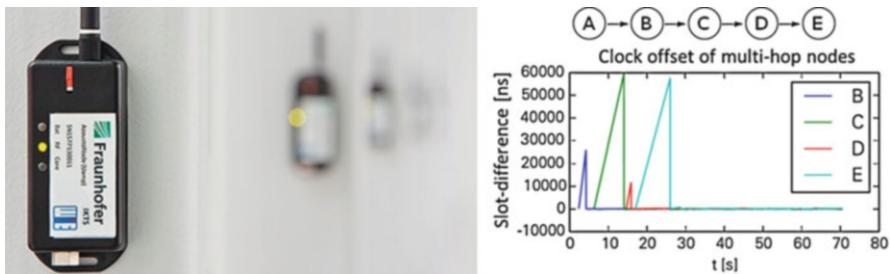
Examples of commercially available WSNs for industrial applications as of 2018:

- LORD Microstrain: WSN based upon a proprietary LXRS protocol with time synchronization of 32 µs in a star topology. Offers sensor nodes for a variety of measurement tasks (acceleration, pressure, strain, temperature, vibration, torque, analog, etc.), [www.microstrain.com](http://www.microstrain.com)
- Space Wireless: WSN based on UWB with a time synchronization of up to 1 µs and 6 months of operation. Offers sensor nodes for a variety of measurement tasks (vibration, shock, analog, temperature, etc.), <http://beanair.com>
- SmartMesh IP: WSN based on TSCH with mesh network topology node synchronization, <http://www.analog.com>
- Permasense: WSN based on WirelessHART for wall thickness measurement of pipes, [www.permasense.com](http://www.permasense.com)

Well-known operating systems for WSN as of 2018:

- TinyOS: Event-based real-time operating system (RTOS) for low-power applications
- Contiki: Event-based RTOS with extensive programming libraries and complete network stacks
- RIOTOS: OpenWSN compatible RTOS with completely implemented IoT-Stack

Beside the mentioned commercial solutions, many other academic prototypes are currently in development for complete SHM solution for bridges and industrial sites (e.g., Fraunhofer IKTS, Fig. 11).



**Fig. 11** Sensor nodes by Fraunhofer IKTS for acousto-ultrasonic, acoustic emission testing, and vibration monitoring with improved time synchronization up to 250 ns per hop, based on OpenWSN

## Summary

In contrast to traditional NDT measurements, SHM measurements demand for cost-sensitive solution since some equipment is permanently mounted to the structure, well-planned concepts for data transfer and data acquisition over longer period of times and long-term stable measurement techniques regarding reliability and repeatability of the measurements to avoid false alarms. Despite these challenges, SHM measurements provide good opportunities and enable the operator of infrastructure and relevant energy installations to get information on their current state regarding integrity and risk assessment. For difficult to access structures like wind turbines or highly safety relevant structures like airplanes, it is an especially useful tool in attempting to reduce risks and furthermore a step toward condition-based maintenance for cost reduction.

## Cross-References

- ▶ [Acoustic Emission](#)
- ▶ [Eddy Current Testing](#)
- ▶ [Guided Wave Testing](#)
- ▶ [High-Frequency Eddy Current Techniques](#)
- ▶ [Optical Fiber Methods in Nondestructive Evaluation](#)

## References

- Bachmann A, Luber M, Poisel H, Ziemann O (2008) Strain sensor using phase measurement techniques in polymer optical fibers. In: Proceedings of the 19th international conference on optical fibre sensors. SPIE, Perth

- Bell GEC, Moore CG (2007) Development and application of ductile iron pipe electrical resistance probes for monitoring underground external pipeline corrosion. Recent development and field experience in corrosion and Erosion monitoring. Dresden, Germany
- Buethe I, Dominguez N, Jung H, Fritzen C-P, Ségur D, Reverdy F (2016) Path-based MAPOD using numerical simulations. In: Smart intelligent aircraft structures (SARISTU), proceedings of the final project conference. Springer, Heidelberg, pp 631–642
- Croxford AJ, Moll J, Wilcox PD, Michaels JE (2010) Efficient temperature compensation strategies for guided wave structural health monitoring. *Ultrasonics* 50:517–528
- Frankenstein B, Fischer D, Weihnacht B, Rieske R (2012) Lightning safe rotor blade monitoring using an optical power supply for ultrasonic techniques. In: 6th European workshop on structural health monitoring. Dresden, Germany
- García-Gómez J et al (2018) Smart sound processing for defect sizing in pipelines using EMAT actuator based multi-frequency lamb waves. *Sensors* 18(3):E802
- Giurgiutiu V (2008) Structural health monitoring with piezoelectric wafer active sensors. Academic, London, ISBN: 9780120887606
- Goldsmith A (2005) Wireless communications, 1st edn. Cambridge University Press, New York, ISBN: 0521837162
- Gratton DA (2013) The handbook of personal area networking technologies and protocols. Cambridge University Press, Cambridge, ISBN: 0521197260
- Hersent O, Boswarthick D, Elloumi O (2012) The internet of things: key applications and protocols, 2nd edn. Wiley, Chichester, ISBN: 1119994357
- Heuer H, Schulze M, Pooch M, Gaebler S, Nocke A, Bardl G, Cherif C, Klein M, Kupke R, Vetter R, Lenz F, Kliem M, Buelow C, Goyvaerts J, Mayer T, Petrenz S (2015) Review on quality assurance along the CFRP value chain – non-destructive testing of fabrics, preforms and CFRP by HF radio wave techniques. *Composites B* 77:494–501
- Hoenig U, Holder U, Pietzsch A, Schulze E, Frankenstein B, Schubert L (2016) Definition of requirements for reference experiments to determine and evaluate various damage mechanisms in fibre composites by acoustic emission. In: 19th world conference on non-destructive testing ISO 17359: Condition monitoring and diagnostics of machines – general guidelines. Beuth
- Kuorilehto M, Kohvakka M, Suhonen J, Hämäläinen P, Hännikäinen M, Hämäläinen TD (2007) Ultra low energy wireless sensor networks in practice. Wiley, Chichester/England/Hoboken, ISBN: 0470516801
- Lovstad A, Cawley P (2011) The reflection of the fundamental torsional guided wave from multiple circular holes in pipes. *NDT&E Int* 44:553–562
- Pankoke S, Friedmann H, Ebert C, Frankenstein B, Schubert L (2011) Structural health monitoring of rotor blades – damage detection by acoustic emission, acoustic ultrasonic and modal analysis. In: 3rd technical conference wind turbine Rotor Blades. Essen, Germany
- Roellig M, Schubert F, Lautenschlaeger G, Boehme B, Franke M, Muench S, Meyendorf N (2012) Technology, functionality and reliability of integrated ultrasonic microsystems for SHM in CFRP airplane structures. In: 4th international symposium on NDT in aerospace. Augsburg, Germany,
- Schubert L, Lieske U, Koehler B, Frankenstein B (2009) Interaction of lamb waves with impact damaged CFRP's – effects and conclusions for Acousto-ultrasonic applications. In: 7th international workshop on structural health monitoring. Stanford, USA
- Schubert L, Lieske U, Koehler B, Frankenstein B (2010) Interaction of lamb waves with impact damaged CFRP's studied by laser-Vibrometry and Acousto ultrasonic. In: 18th European conference on fracture. Dresden, Germany
- Tscheliesnig P, Lackner G, Jagenbrein A (2016) Corrosion detection by means of acoustic emission (AE) monitoring. In: 19th world conference on non-destructive testing. Munich, Germany
- Zhang S, Fu F (2011) Piezoelectric materials for high temperature sensors. *J Am Ceram Soc* 94(10):3153–3170

---

# Index

## A

- Abrikosov flux lines, 1271  
Absolute defect response amplitude (ADRA), 802  
Absolute defect response signal (ADRS), 798  
Achromatic phase shifter, 434  
Acoustic Emission, 620–621, 1558  
    advantages and limitations of, 137  
    data analysis, 119–123  
    future perspectives, 136  
    hits distribution, 119–120  
    in monitoring concrete structures, 118  
    monitoring of damage and repair of self-healing concrete, 123–135  
    principle, 116  
    signal complexity problem, 135  
    source localization, 120  
    testing, 1594–1596  
    wave features, 122  
Acoustic nonlinearity, 253  
Acoustic waves  
    propagation, 252  
    surface, 175  
Acousto-elastic effect  
    basics, 236–238  
    definition, 236  
Acousto-optic modulator, 342  
Active thermography  
    classical methods of, 1406  
    energy sources, 1406–1408  
    function and design, 1402  
    lock-in thermography, 1412–1417  
    pulse thermography and pulsed phase thermography, 1408–1412  
    thermographic methods, 1402–1406  
Advanced eddy-current testing and evaluation,  
    *see* Eddy-current testing (ECT)  
Aero engine components, 747, 748  
Air coil sensors, 833  
Air-coupled transducers, 159  
Algebraic reconstruction technique (ART), 1177, 1254–1257  
Alternative thermal wave method (ATWM), 1444  
Aluminum, barrier film on, 355  
Aluminum gallium arsenide (AlGaAs), 1392  
Aluminum plate bulk waves  
    baseline subtraction, 67  
    energy accumulation, 69  
    experimental configuration, 65  
    phase velocity filtering, 67  
    radial B-scans, 68  
    Rayleigh wave, 66  
    wavefield snapshots, 66  
American Society for Nondestructive Testing (ASNT), 648  
3-Amino-5-mercapto-1,2,4-triazole (AMTA), 589  
Amorphous polymer, 1336–1338  
Ampère's Law, 992  
Analyzer-based imaging (ABI)  
    applications of, 1061  
    fundamentals, 1058–1061  
Anisotropic magneto-resistance (AMR)  
    magnetometers, 836  
    sensors, 681  
Antenna, 937  
Apparent eddy current conductivity (AECC), 731, 744  
Argon-ion lasers, 341  
Artificial intelligent approach, 718  
Atomic nonlinearity, 298  
Attenuated total reflection infrared spectroscopy (ATR-IR), 911  
ATWM, *see* Alternative thermal wave method (ATWM)

- Austenitic stainless steel, 860, 862  
 Automatic fringe pattern analysis, 365  
 Axial birefringence, 535–536
- B**  
 Barrier film, on aluminum, 355  
 Basic Law of Photometry, 1381–1382  
 Beam hardening, 1105, 1108  
 Beamsplitter, 520  
 Beer's law, 1055  
 Berlin research reactor, 1244  
 Besselogram, 372  
 Biaxial stress, 556–557  
 Bidimensional empirical mode decomposition (BEMD), 368  
 Bimaterial interface, bulk waves  
     general solution, 20–23  
     P-wave excitation, 23–25  
     solid-solid interface, 17–20  
     SV-wave excitation, 25–28  
 Biofouling, in water processing, 481–482  
 Biological samples, inspection of, 486, 487  
 Biot-Savart's law, 651, 655, 697, 698  
 Blackbody radiation, 1377  
 Blade fatigue, 1485  
 BLS, *see* Brillouin light scattering (BLS)  
 Bolometeter detector, 1506  
 Boltzmann constant, 1374  
 Bonded plate guided waves, 61  
 Boundary based methods, 1131  
 Boundary conditions, 1359–1360  
 Boundary element method (BEM), 666, 668, 671, 672  
 Bragg edge imaging, 1218–1295  
 Breast microwave radar (BMR), 760  
 Bremsstrahlung, 1170  
 Brillouin-based distributed sensor systems, 613  
 Brillouin light scattering (BLS), 328  
 Brillouin scattering, 329  
     mechanism, 329–340  
     surface ripple-mediated, 331  
     transparent solids, 329  
 Brittle-ductile transition temperature (BDTT), 576  
 Broadband source, 423  
 Bulk waves  
     bimaterial interface, 16  
     infinite elastic medium, 4–16  
     liquid and solid media, 4–28  
     in liquids, 28–30

- C**  
 Carbon composites  
     basic principles, 733–735  
     characterization of defects, 737, 739  
     texture and waviness, 735, 738  
 Carbon fiber-reinforced plastics (CFRP), 908, 1417, 1597  
     composites, 1575  
 Carbon fiber-reinforced polymers (CFRP), 287, 1071, 1076, 1515  
 Carrier-frequency spatial phase shift  
     shearography, 403  
 Cavity perturbation, 952  
 Ceramic-matrix composites (CMC), 740, 741  
 Charge-coupled device (CCD)  
     camera, 1553  
     detector, 563  
 CL, *see* Computed laminography (CL)  
 Classical nonlinearity, 297–298  
 Classification methods, 1539  
 Clustering technique (K-means algorithm), 1043  
 Coated solid, surface Green's function for, 335  
 Coatings/paints, 491  
 Coating thickness measurement, 978  
 Coherence condition, solid-solid interface, 19–20  
 Coherence radar, *see* Fourier domain OCT (FD OCT)  
 Coherence scanning microscopy, 470  
 Coherence time, 425  
 Coherent optical time domain reflectometry (C-OTDR), 634  
 Coincidence Doppler broadening (CDB), 1322  
 Complete electrode model (CEM), 760  
 Complex degree of coherence, 428  
 Complex refractive index, 1056, 1057  
 Composite(s), 1597–1600  
     materials, 485, 486  
     panel guided waves, 57  
 Compressive stress, 567  
 Computed laminography (CL), 1113  
 Computed tomography (CT), 1174  
 Compute unified device architecture (CUDA), 435  
 Conductivity, ECT  
     alloys, 663  
     conductive coating, 664  
     hardness, 664  
     temperature and residual stresses, 664  
 Cone beam geometry, 1099

- Conjugate gradient (CG) method, 703, 707, 709, 715, 717, 721  
Constant phase element (CPE), 902  
Contact-based sensors, 1547  
Continuous-wave (CW) radar, 1001  
Contrast limited adaptive histogram equalization (CLAHE) method, 1030  
Contrast-transfer function (CTF), 1066  
Correlogram, 432  
Corrosion, 627–628  
    defect in pipeline, 843  
    inspection, 491  
Corrosion process, Raman spectroscopy in, 585  
Cosmic ray events (CREs), 563  
CPE, *see* Constant phase element (CPE)  
Crack  
    depth, 1509–1510  
    detection under non-conducting coatings, 1511  
    width, 1510  
Crack reconstruction, ECT signals, 707  
    artificial intelligence approach, 718  
    metaheuristic methods, parallel tabu search, 720  
    multiple cracks, 721  
    planar crack, 708  
    SCC, 712  
Cross-field effect, 836  
Crystal disorder, 566  
Crystalline silicon, 349–350  
Cubic boron nitride (cBN), 551, 552
- D**
- Dark-field contrast intensity, 1289  
Data processing algorithms, 1368  
DEFECTOVISION® IR testing system, 1513, 1514  
Defect response signal (DRS), 805  
Diamond anvil cell (DAC), 555  
Diamond-like carbon (DLC) films, 357  
DIC, *see* Digital image correlation (DIC)  
Dielectrics  
    reflection test, 940  
    transmission test, 944  
Differential Lorentz force eddy current testing (DiLET), 800–801  
Diffraction, 366, 375  
Diffraction-enhance imaging (DEI), 1059, 1061  
Diffuse scatter  
    Born approximation, 1193  
    semiconductor, 1196  
topological roughness, 1192  
Yoneda wings, 1196–1197  
Digital image correlation (DIC), 1548–1550  
    in aerospace applications, 1550–1552  
    in civil applications, 1552–1562  
    in dynamic applications, 1562–1565  
    in energy applications, 1566–1572  
    in material testing applications, 1572–1577  
Digital image processing, 1028  
Digital magnetometers, 836  
Digital shearography  
    fringe formation, 388–390  
    fringe interpretation, 393  
    image sensors, 385  
    laser device, 405  
    for NDT, 404–414  
    relative phase change and deformation, 390–393  
    schematic setups, 386–388  
    shearing device, 405  
    spatial phase shift shearography, 398–404  
    temporal phase shift, 393–398  
Dilatation wave equation, 9  
3 Dimensional ECT analysis  
     $A-\phi$  method, 667  
     $Ar$  method, 673  
Diode-Pumped Solid State (DPSS) lasers, 341  
Direct current (DC) stress, 621  
Dirichlet boundary conditions, 792  
Discrete cosine transform, 1535  
Discrete Fourier transform, 1535  
Discrete time signal, 1527  
Discrete wavelet transform, 1535, 1537  
Discretization  
    boundary elements, in air region, 671  
    finite element, in conductor region, 668  
Discriminant analysis, 1019  
Dispersion diagram for aluminum plate, 145  
Displays, 495, 496  
Distortional waves, 11  
Distributed acoustic sensing (DAS), 615  
Doppler broadening spectroscopy (DBS)  
    CDB, 1322–1323  
    data analysis, 1320–1322  
    experimental setup, 1318–1320  
Double D coils, 1506  
3D Raman mapping, 566–584  
Dynamic stressing, 412

**E**

ECT probes, *see* Eddy-current testing (ECT)  
 Eddy current magneto-optic imaging (MOI) sensors, 839  
 Eddy-current testing (ECT), 646, 659, 706, 782, 786  
   advantages, 650  
 $A - \phi$  method, 667  
 AMR sensor, 681  
 $Ar$  method, 673  
 basic formulation, of fast ECT simulation scheme, 704  
 basic principles of, 649  
 conductivity, 663  
 crack reconstruction, 707  
 decreasing test frequency, 660  
 deep crack detection, 687  
 discontinuities, 666  
 eddy current perturbation, 695  
 electromagnetic field, basic  
   equations of, 651  
 excitation field strength, increase of, 661  
 fluxgate sensors, 679  
 frequency, 664  
 geometry, 665  
 GMR sensor, 682  
 Hall sensors, 680  
 induction coil, 676  
 and LET, 788–790  
 limitations, 650  
 limiting factor of, 787  
 magnetic field measurement, crack detection, 689  
 material properties, 662  
 optimal design, crack detection, 698  
 permeability, 664  
 phenomenological relationship, 697  
 pick-up signals, formula for, 706  
 probe handling, 665  
 proximity, 664  
 skin effect, standard depth, 657  
 source magnetic field and induced eddy currents, 691  
 SQUID, 679  
 surface crack testing, absolute pancake probe for, 684  
 TMR sensor, 682  
 transmitter and receiver technique, 661  
 tube inspection, 686  
 unflawed field information, databases of, 705  
 Edge/end effect, 665  
 Edge illumination (EI), 1079

EIS, *see* Electrochemical impedance spectroscopy (EIS)  
 Elastic constants, 236  
   determination, 353–357  
 Elastodynamic finite integration technique (EFIT), 78, 163  
 Elasto-optic scattering, 329  
   in transparent media, 329–330  
 Electrical conductivity, 663  
 Electrical impedance spectroscopy (EIS), 492  
 Electrical resistance, 1596  
 Electrical resistance tomography (ERT), 759, 760  
 Electrical resistivity, 761, 763  
 Electric field, 654  
 Electrochemical impedance spectroscopy (EIS)  
   atmospheric corrosion, 902–905  
   attenuated total reflection infrared spectroscopy, 911  
   Bode plot and Nyquist plot, 909  
   corrosion in concrete, 905–908  
   corrosion rates, 903  
   corrosion studies, 900  
   electrode systems, 901  
   facile technique, 908  
   high-temperature applications, 915–916  
   hot corrosion monitoring, 916  
   inhibitor film persistency, 913  
   inhibitors, 913–915  
   instrumentation, 901–902  
   local electrochemical impedance mapping, 913  
   organic coatings, 910  
   rebar corrosion, 905  
   soil corrosion, 916  
   spectra, 911  
   steel surfaces, 907  
   water absorption by polymer coatings, 910  
   weight loss measurement, 905  
 Electromagnetic(s)  
   for ECT Problem (*see* Eddy-current testing (ECT))  
   skin depth, 1499, 1502  
   wave propagation, 992  
 Electromagnetic acoustic transducers (EMAT), 149, 159  
 Electromagnetic compatibility (EMC), 1001, 1015  
 Electromagnetic testing (ET), 648  
 Electron discharge machining (EDM), 577  
 Electronic shearography, 385  
 Electronic waveguide systems, 981

- Electro-optics (EO) polarization rotator, 525, 526  
EL2-induced defect, 1334–1336  
Ellipsometry  
    axial birefringence, 535  
    definition, 515  
    imaging microellipsometer, 517–521  
    index of refraction variation, 530–531  
    microellipsometer with rotational symmetry, 521–525  
    micro-optical components, 536–537  
    nulling microellipsometer, 525–529  
    objective lenses, 516  
    sub-micron PMMA lines measurement, 534  
    surface topology, 529–530  
Embedded copper nanoprecipitate, 1329–1332  
Embedded piezoelectrics  
    integrated castAl/PI/PZT piezomodules, 1461–1465  
    LTCC/PZT sensor-actuator modules (see LTCC/PZT sensor-actuator modules)  
        macro-fiber composite actuators, 1455–1458  
    thermoplastics-compatible piezoceramic modules, 1457–1461  
Emissivity, of real bodies, 1381  
Empirical mode decomposition (EMD), 368  
Endpoint memory, 298  
Equivolume waves, 11  
EVITA project, 1076  
Extended area approach (EAA), 805  
Extrinsic FPI (EFPI) sensor, 609, 610, 623, 626
- F**  
Fabry-Pérot interferometry, 329  
Fabry-Perot sensors, 608–611  
    EFPI sensor, 609, 610  
    high-performance materials, 622–625  
    interference function of, 609  
    types of, 608  
Factorization method (FM), 759  
Fan beam X-ray computed tomography (XCT), 1102  
Faraday's law, 650, 652  
Fast Fourier Transformation (FFT), 260  
Fatigue cracks, sonic IR detection of, 1485  
Fatigue process, in polycrystalline diamond, 577  
FDK algorithm, 1101  
Feature extraction methods, 1534  
Feature reduction methods, 1538
- Ferrite core transducer, 1498  
Ferritic steel, 1513  
Ferroelectric crystals, 500, 501  
Ferromagnetic metals, 1502  
Fiber Bragg grating (FBG) sensors  
    Bragg wavelength, 602, 605  
    coatings of, 607  
    commercial FBG pressure sensors, 608  
    commercial FBG temperature sensors, 607  
    electromagnetic immunity, 607  
    FBG array, 605  
    pre-fabricated, 601  
    for static and low-frequency strain changes, 615–617  
    for strain-based dynamic responses, 617–620  
    TDM, 605  
    tunable/swept-wavelength light sources, 604  
Fiber optic sensors, 1547  
Fiber-reinforced polymer (FRP) RC beams, 1575  
Field cooling, 1260  
Fill factor, 665  
Filtered back projection, 1280  
Filtering frequency wavenumber, 56–58  
Fingerprint identification, 483, 484  
Finite difference simulations, 163  
Finite difference method (FDM), 666  
Finite element analysis (FEA), 632, 793  
Finite element discretization, in conductor region, 668  
Finite element method (FEM), 655, 666–668, 671, 695  
    of complicated cracks, 706  
    with edge elements, 675  
Finite element model (FEM), 1557  
Finite impulse response (FIR), 1014  
Fixed grid methods, 791  
Fizeau interferometer, 456  
Flip-efficiency, 1249  
Flow characterization, 479, 480  
Fluxgate sensors, 679, 835–836  
Focal plane arrays, 1382  
Focused beam rigorous coupled wave analysis (FB-RCWA), 533  
Fourier domain OCT (FD OCT), 473  
Fourier transform (FT), 367, 1066, 1078, 1232–1235  
Four-point-calibration, 731  
Frequency domain imaging (OFDI), 473  
Frequency domain OCT, *see* Fourier domain OCT (FD OCT)

- Frequency-modulated continuous wave (FMCW), 968, 971
- Frequency-modulated interrupted continuous-wave (FMICW), 1002
- Frequency-wavenumber analysis
- Hanning window, 54–55
  - slices, 55–56
  - spatial sampling, 53
  - temporal sampling, 53
  - wavelength elongation, 55
- Fresnel diffraction, 1061, 1062
- Friction stir welding (FSW), 811–812
- imperfections/defects, 814
  - MIECT measurements of, 815–819
  - nondestructive testing of, 814–815
- Fringe analysis, white light interferometry
- sampling, light interference signal, 435
  - scanning and vibration error
    - self-calibration, 436
    - signal processing, 432–435
- Full-field OCT, 471
- Full-wave simulation, 947
- Full width at half height maximum (FWHM), 1194–1195
- Functionally graded materials (FGM), 1572
- Functional OCT-imaging
- flow characterization, 479–481
  - polarization-sensitive OCT (PSOCT), 478–479
- FWHM, *see* Full width at half height maximum (FWHM)
- 
- G**
- Gallium arsenide (GaAs), 1392
- Gamma coherence function, 425
- Gauss function, 1226, 1229
- Gauss' law, 652, 653
- electric field, 992
  - magnetic field, 992
- Gauss-meter, 948
- Geometric optics model, 517
- Giant magneto-resistance (GMR)
- magnetometers, 837
  - sensor, 682
- Glass fiber-reinforced plastics (GFRP), 1599
- Glass fiber-reinforced polymers (GFRP), 287
- Glass laminate aluminum reinforced epoxy (GLARE), 807–811
- Goal function scanning (GFS) method, 811
- Gouted joints, 1602
- Gradient-index optics, 500
- Grain structure, 97–99
- Graphical user interface (GUI), 86
- Grating-based phase-contrast imaging (GB PCI)
- cultural heritage, 1077
  - fundamentals, 1071–1075
  - industrial inspection and security scanning, 1076
  - materials, 1076
  - tissue engineering and related products, 1076–1077
  - wavefront sensing for optics metrology, 1076
- Grating interferometry, 1289–1292
- Grazing incidence reflectivity, *see* X-ray reflectivity (XRR)
- Green's function, 329
- surface, 334
- Ground penetrating radars (GPRs), 854
- antennas, 996–999
  - bistatic, 991
  - conduction, 995
  - data processing, 1005–1010
  - electrical parameters of materials, 996
  - electrical properties, 995
  - electromagnetic wave propagation, 992–994
  - features, 990
  - frequency-domain, 991
  - kernel based modeling, 1017, 1018
  - monostatic, 991
  - multistatic, 991
  - operating principle, 989
  - polarization, 995
  - PSTs (*see* Postprocessing support tools (PSTs))
  - software, 1011
  - system specification, 999–1004
  - time-domain/pulsed radars, 991
- Guided ultrasonic waves (GUW), 1558
- Guided waves (GW), 1593
- Guided wave testing
- anisotropic structures, composites, 153–156
  - curved structures, 156
  - defect detection in isotropic (metal) plates, 147–151
  - experimental methods, 158–162
  - multi-layered plates, stiffeners, and lap-joints, 151–153
  - pipes, 156–158
  - propagation, 143–147
  - simulations, 162–163

**H**

Haar transform, 1535  
Hagen-Rubens relation, 1379  
Halbach systems, 802  
Hall effect, 680  
Hall sensors, 834  
Hanger cables, 1557–1558  
Hard vs. soft computing, 1531  
Harmonic generation, 253, 302  
Harmonic waves, 8, 30  
Heat conduction, differential equation of, 1361  
Heat transfer, 1357–1361  
in defects, 1360–1361  
Helmholtz coils, 1505  
Helmholtz decomposition concept, 8  
Hertzian friction, 300  
Hertzian impact theory, 275  
Hidden Markov model (HMM), 1020  
Higher harmonic generation, 253  
High frequency eddy current (HFEC) technique  
basic principle, 730–731  
carbon composites, 733–735  
ceramic composites, 740–741  
characterization of defects, 737–740  
instrumentation, 731–732  
low electrical conductivity, 741–754  
polymers and dielectrics, 752–755  
texture and waviness, 735–738  
High-frequency ultrasonic probe, 86  
Hilbert-Huang transform (HHT) based fringe analysis, 368  
Hilbert spiral transform, 369  
Holotomography, 1063, 1066  
Hooke's law, 297  
Hughes, D.E., 647  
Hydrated polymeric oxide model, 588  
Hydrostatic piezo-Raman coefficient, 553–555  
Hyper-Raman effect, 546  
Hysteros, 299

**I**

iDAS, 615  
IE, *see* Impact-echo (IE) technique  
Imaging microellipsometer  
instrumentation, 520–521  
working principle, 517–520  
Impact-echo (IE) technique, 273  
Induction coil, 676  
Induction current distribution, 1500  
Induction thermography  
applications in steel industry, 1498  
carbon fiber-reinforced polymers, 1515  
defect detection, 1507–1511  
electromagnetic and thermal key parameters, 1499–1500  
hidden defects in ferritic steel, 1513–1514  
inductors, 1505  
infrared camera, 1506  
instrumentation, 1504–1505  
vs. magnetic particle testing, 1519  
material parameters, characteristic lengths and induction heating efficiency, 1500–1503  
pre-processing of the recorded image sequences, 1506  
rail surface defects, 1516  
scheme of, 1498  
semiconductors, 1515  
standardization, 1520  
steel components, 1511–1513  
wheel surface defects, 1517  
Inductive heating, 1407  
In-elastically scattered light, 328  
Infinite elastic medium, bulk waves  
dilatational waves, 9  
irrotational and equivolume waves, 10–12  
plane waves, 4–6  
types, 6–8  
wave potentials, 8–9  
z-invariant waves, 12–16  
Infrared cameras, 1394  
Infrared characteristics, of bodies, 1381  
Infrared thermographic testing (ITT), 1350  
analytic solutions, 1362  
inspection procedure, 1351–1355  
relationship in, 1362–1364  
terminology, 1350–1354  
Inline inspection system, 982  
In-plane displacement and strain analysis, 376  
Integrated cast Al/PI/PZT piezomodules, 1461  
Intelligent signal processing, 1529, 1530  
Interface conditions, solid-solid interface, 18  
potentials, 21–22  
wave angles and material properties, 22

- Interference  
 encoding information, 363  
 fringes, 424  
 information decoding by automatic fringe pattern analysis, 365–370  
 Moiré interferometry, 374  
 vibration testing, 370
- Interferometry, 1058
- Ion-implantation-induced defect, 1333–1336
- Iron atoms oxidation, 585
- Iterative closest point (ICP), 1568
- ITT, *see* Infrared thermographic testing (ITT)
- J**  
 Josephson effect, 679
- K**  
 Kaczmarz algorithm, 1255  
 Karhunen-Loeve transform, 1535  
 Kirchhoff's Law of Radiation, 1376
- L**  
 Lagrange multiplier approach, 792  
 Lambert-Beer law, 1170  
 Lamb waves, 153  
 Langmuir-Blodgett films, 1183  
 Laplace equation, 671  
 Laplacian operator, 654  
 LARS, *see* Local acoustic resonance spectroscopy (LARS)  
 Laser(s), 341  
 acoustic devices, 173  
 structuring and welding, 493  
 thermography, 1408  
 vibrometer, 161  
 Laser Doppler vibrometer (LDV), 44  
 Laser generation, of surface acoustic waves, 184  
 Laser-hardened steels, 219–225  
 Laser intensity modulation method (LIMM)  
 deconvolution technique, 1442  
 polynomial thermodynamic potential, 1439  
 spatial resolution, 1440–1442  
 two-dimensional, 1443  
 Lateral wave or guided mode (LGM), 334  
 Lattice crystal defect, 1303–1304  
 LDV, *see* Laser Doppler vibrometer (LDV)  
 Leakage field method, 855  
 Lenz's law, 650  
 Levenberg-Marquardt-procedure, 193
- Lift-off, 665  
 Light-emitting diodes (LEDs), 495, 496  
 LIMM, *see* Laser intensity modulation method (LIMM)  
 Linear attenuation coefficient, 1055, 1056, 1066  
 Linear sampling method (LSM), 759  
 Line spread function (LSF), 1226  
 Linnik interferometer, 439  
 Liquid penetrant testing, 1519  
 Liquids, bulk waves  
 stresses, 30  
 wave motion, 28–30  
 Liquid-solid (LS) interface  
 boundary conditions at, 31  
 bulk waves at, 32–35  
 cases, 31  
 coherence condition, 33–34  
 critical angles, 34  
 displacements and stresses, 32  
 interface conditions, 33  
 Local acoustic resonance spectroscopy (LARS), 273  
 applications, 274  
 automated ball drops, 281–282  
 body/body collisions, 275–276  
 carbon fiber-reinforced polymers, 287–290  
 emitter/impactor, 279–280  
 glass fiber-reinforced polymers, 287  
 impactor stiffness and force, 277–278  
 materials testing, 276  
 modal analysis hammer, 281  
 receiver/microphone, 280  
 signal processing, 282–284  
 wind turbine blades, 284  
 Localized fringes, 423  
 Localized phased array type transducers, 150  
 Lock-in thermography, 1412–1417  
 Logical expression approaches (LEA), 793–794  
 MDA, 796  
 MMA, 795  
 QSA (*see* Quasi-static approach (QSA))  
 Logical expressions (LE), 793, 794  
 Longitudinal guided mode (LGM)  
 velocity, 335  
 Long-period grating (LPG) sensors, 639  
 Lorentz force eddy current testing (LET), 782  
 benchmark problem, 793  
 DiLET, 800  
 and ECT, 788  
 general principle of, 785  
 GLARE (*see* Glass laminate aluminum reinforced epoxy (GLARE))

- LEA (*see* Logical expression approaches (LEA))  
permanent magnet system, optimization of, 801–803  
theory, 786–788  
working principle of, 783
- Lorentz force evaluation (LFE), 786, 803–806
- Lorentz force sigmometry (LoFoS), 785
- Lorentz force velocimetry (LFV), 785
- Lorentz function, 1229
- Low field nondestructive evaluation  
applications, 841–842  
buried pipelines and structures, 854–857  
clinical applications, 873  
composite and sandwich structures, 861–866  
crack detection and profiling, 860–861  
eddy current magneto-optic imaging  
sensors, 839  
fluxgate, 835–836  
Hall sensors, 834–835  
induction coils, 830–834  
mechanical properties and stress, 867–871  
metrology, 873–874  
multilayer joints, 865–867  
permanent magnet defects, 852–854  
pipeline and tube defects inspection, 840–849  
rail wheels defects, 870–872  
SQUID, 838–839  
subsurface inspection, 849–853  
wire rope damages, 855–860
- LTCC/PZT sensor-actuator modules  
embedding piezoelectric slab, 1452  
Nyquist plots, 1454  
schematic design, 1451  
transmission thermal wave microscopy, 1454–1455
- Lumped model, 1432
- M**
- 3MA approach, 887
- Mach-Zehnder interferometer, 401–403
- Macro-fiber composite actuators, 1455
- Magnetic Barkhausen noise (MBN) analysis, 884–885, 890
- Magnetic crystal Bragg diffraction, 1245
- Magnetic domains, 883, 1258–1260
- Magnetic flux density, 773
- Magnetic flux leakage (MFL) detection, 820, 840, 843, 860
- Magnetic hysteresis, 883
- Magnetic materials, 883
- Magnetic particle testing, 1519
- Magnetic permeability, 652, 664
- Magnetism, 883
- Magnetoencephalography (MEG), 874
- Magnetoresistance magnetometers, 836  
anisotropic magnetoresistance, 836  
giant magnetoresistance, 837  
tunneling magnetoresistance, 837
- Markov models, 1020
- MAR, *see* Metal artifact reduction (MAR)
- Material machining and processing, 493, 494
- Maxwell-Ampere's law, 652
- Maxwell's equations, 652, 656, 993
- Mean square error (MSE), 708
- Mean-square residual, 708
- Mechanical impedance analysis (MIA), 273
- Meissner–Ochsenfeld effect, 1260
- Mesoscopic friction models, 301
- Mesoscopic structure, 299
- Metal artifact reduction, 1109
- Metal matrix composites (MMCs), 1067
- Metrology, 873
- MIA, *see* Mechanical impedance analysis (MIA)
- Michelson interferometer, 403, 437, 470
- Microbiologically influenced corrosion (MIC), 916
- Micro-CT ( $\mu$ -CT), 1175
- Microellipsometer with rotational symmetry  
instrumentation, 524–525  
working principle, 521–524
- Micromagnetic materials, multi-parameter  
approaches in, 892–895
- Micromagnetic microscopy, multi-parameter  
advantage in, 894–895
- Micromagnetics, 882–896  
calibration procedure, 887–889  
eddy current impedance analysis, 886  
hybrid approach, 887  
incremental permeability analysis, 885–886  
pattern recognition, 889  
regression analysis, 888–889  
validation procedures, 889–892
- Microwave and millimeter wave NDE, 931–934  
energy associated with microwaves, 934–935  
instrumentation, 938–939  
microwave microscopy, 945–950  
radiometry, 959–960  
reflection test, 940–942

- Microwave and millimeter wave NDE (*cont.*)  
 resonant testing methods, 950–957  
 scattering tests, 943–945  
 theory, 935–938  
 transmission tests, 942–943  
 ultra wide band radar, 957–959
- Microwaves, 1407
- Minimum quadratic difference (MQD) algorithm, 1565
- Mirau interferometer, 437
- Modal analysis hammer, 281–282
- Mode conversion, 26
- Modulation transfer function (MTF), 1229
- Moiré interferometry, 374–377
- Monotonicity principle, in Eddy current tomography  
 imaging method, 762–764  
 in large skin-depth regime, 761–762  
 pulsed eddy current imaging method, 772–775  
 sign index, 764  
 small skin-depth regime, 769–772
- Motion-induced eddy current testing (MIECT), 782  
 ferromagnetic materials, 819–821  
 force measurement, 799–800  
 FSW (*see* Friction stir welding (FSW))  
 LET (*see* Lorentz force eddy current testing (LET))  
 LFE, 804  
 material science, 819  
 nondestructive defect detection, 819  
 process control and monitoring, 819
- Motion-induced eddy current testing and evaluation (MIECTE), 805
- Moving defect approach (MDA), 795–796, 798
- Moving grid methods, 792
- Moving magnet approach (MMA), 794–795, 798
- Mueller matrix polarimetry (MMP), 538
- Multi-channel spatial phase shift shearography, 403
- Multi-frequency eddy current (MFEC), 648
- Multipass Fabry-Pérot interferometry, 329
- Multipass tandem Fabry-Pérot (FP) interferometer, 341–342
- Multiple cracks reconstruction, 721
- Multiple input-multiple output (MIMO) array antenna technology, 1004
- Multiple input-multiple output (MIMO) method, 974
- Multiple quantum wells (MQW), 1392
- Multivariate curve resolution with alternating least squares (MCR-ALS), 585
- Mutual induction probe, 678
- N**
- Natural materials, inspection of, 486, 487
- Navier-Lame equations, 8, 9
- Neutron(s)  
 index of refraction, 1219–1222  
 interactions with matter, 1222–1223  
 polarization, 1252–1254  
 properties, 1220
- Neutron radiography  
 beam polarization, 1248  
 geometries used in, 1224–1226  
 interaction with magnetic field, 1249–1251  
 magnetic field calculation, 1251–1252  
 mathematics of computerized tomography, 1230–1232  
 phase-based imaging signals, 1274–1288  
 polarizing neutrons, 1245–1248  
 projection-slice theorem, 1232–1236  
 reconstruction of continuous magnetic fields, 1260  
 reconstruction techniques, 1254–1260  
 spatial resolution, 1226–1229  
 theory of back-projection, 1236–1243
- Neutron tomography  
 beam polarization, 1248  
 computerized tomography, 1230–1232  
 geometries used in, 1224–1226  
 interaction with magnetic field, 1249–1251  
 magnetic field calculation, 1251–1252  
 phase-based imaging signals, 1274–1288  
 polarizing neutrons, 1245–1248  
 projection-slice theorem, 1232–1236  
 reconstruction techniques, 1254–1260  
 spatial resolution, 1226–1229  
 theory of back-projection, 1236–1243
- Nickel-based superalloys, 351
- Noise-equivalent temperature difference (NETD), 1383
- Noise-modulated continuous waveform (NMCW), 1002
- Non classical nonlinearity, 298–301
- Non contact displacement system, 1556
- Nonlinear acoustics  
 definition and historical background, 252–253  
 earlier efforts, 253–256
- Nonlinear hysteretic model, 311–314

- Nonlinear impact resonance acoustic spectroscopy (NIRAS), 255
- Nonlinear laser ultrasonic technique, 258
- Nonlinear resonance technique, 255
- Nonlinear reverberation spectroscopy
- application to thermally loaded CFRP, 305–314
  - description, 304–305
  - fatigue damage in CFRP and steel, 314–318
- Nonlinear ultrasonic bulk waves NDT, 259–261
- Nonlinear ultrasonic guided waves NDT, 260–264
- Nonlinear ultrasonic NDT, 258–264
- Nonlinear waveform modulation spectroscopy, 302–303
- Nonlinear wave modulation spectroscopy, 253–255
- Nonlinearity in sonic IR imaging, 1492
- Nonlocalized fringes, 423
- Nonthermal effects of microwave radiation, 935
- Normalized root-mean-square-error (NRMSE), 811
- N+2 temporal phase shift algorithm, 398
- Nulling ellipsometer, 515
- Nulling microellipsometer
- instrumentation, 528–529
  - working principle, 525–528
- O**
- OCT-based velocimetry, 479, 481
- Oersted's law, 649
- Ohm's law, 786
- One axial stress state evaluation, 238
- Opaque solids
- surface ripple-mediated Brillouin scattering from, 331
- Operating deflection shapes (ODSs), 1567
- Optical-based technique, 1547
- Optical coherence tomography (OCT), 470
- axial resolution, 472
  - Fourier domain OCT, 473
  - full-field OCT, 471
  - light sources, 474
  - Michelson-interferometer, 470, 471
  - scanning, 475
  - spectral domain OCT, 473, 474
  - time domain OCT, 471
  - transversal resolution ( $\Delta x$ ), 472
  - wavelength, 473
- Optical coherence tomography (OCT)
- applications
  - biofouling in water processing, 481, 483
  - biological samples and natural material inspection, 486, 487
  - composite materials, 485, 486
  - ferroelectric crystals, 500, 501
  - functional OCT-imaging, 478–481
  - gradient-index optics, 500
  - nanoparticles and polymers, 476–478
  - optical surface inspection, 498, 499
  - printed electronics, 496, 498
  - thin-film coatings, 487–492
- Optical fiber sensors (OFS), for NDE, 596, 598–599
- Brillouin-based distributed sensor systems, 613–615
- categories, 598
- corrosion, 627
- distributed acoustic sensing, 634–635
- distributed strain measurement, 631–634
- distributed temperature measurement, 631
- FBG sensors (*see* Fiber Bragg grating (FBG) sensors)
- feature of, 612
  - fiber Fabry-Perot sensors (*see* Fabry-Perot sensors)
  - light beam propagation, 600
  - multiplexed single-point and segmented sensors, 611–612
  - opto-electronic unit, 597
  - PD sources, 620
  - Raman-based distributed sensor systems, 613
  - Rayleigh scattering, 612
  - segmented, 628–631
  - sensing unit, 596
  - standardization, 635–638
- Optically based measurement techniques, 1579
- Optical spectrum analyzers (OSA), 603
- Optical surface inspection, 498, 499
- Optical time-domain reflectometry (OTDR), 611, 613
- Optic modes, 328
- P**
- PALS, *see* Positron annihilation lifetime spectroscopy (PALS)
- Parallel-beam synchrotron X-ray computed tomography (XCT), 1103
- Parallel magnetized magnet segments and Hall sensors, 859

- Partial differential equation (PDE), 758  
 Partial discharges (PD), 620  
 PAS, *see* Positron annihilation spectroscopy (PAS)  
 Passive thermographic techniques  
   preventive maintenance, 1397  
   process monitoring, 1400  
   testing electronic components and assemblies, 1398  
 Permanent magnet synchronous motors (PMSMs), 852–854  
 Phase-contrast imaging  
   ABI, 1058–1061  
   edge illumination, 1079–1083  
   GB PCI, 1071–1077  
   interferometry, 1058–1059  
   PB PCI, 1061–1071  
   tracking-based, 1077–1079  
 Phase-contrast XCT  
   propagation based, 1151  
   Talbot Lau interferometer, 1155  
 Phased arrays, 102  
 Phase domain techniques, 366  
 Phase-resolved Doppler-OCT, 480, 481  
 4+1 Phase shift algorithm, 396  
 4+4 Phase shift algorithm, 393  
 Phase shifting interferometry (PSI)  
   techniques, 430  
 Phenomenological approach, 299  
 Photogrammetry, 1548, 1550  
 Photographic shearography, 385  
 Photoluminescence, 542  
 Photometry, basic law of, 1382  
 Photon detector cameras, 1506  
 Photonic-crystal fibers (PCF), 638  
 Photon sensors  
   maximum detectivity, 1390  
   operational principle of, 1389  
   photodiodes, 1391  
   quantum-well sensors, 1391–1392  
   semiconductors, 1390  
   sensors, 1392  
   thermal sensors and, 1393  
 pH sensors, 628, 629  
 Physical vapor deposition (PVD), 201  
 Pick-up coil sensors, 830  
 Piezoelectric transducer (PZT), 260  
 Pipeline and tube defects inspection, 840–849  
 Planar crack reconstruction, ECT signals, 712  
   basic formulation, 708  
   EDM crack, parameterization of, 710  
 Planck's Law of Radiation, 1374  
 Planck's Radiation Law, 1372  
 Plane bulk waves, 4  
 Polarized neutron imaging  
   radon transform for image reconstruction, 1261–1266  
   superconductors type I, 1266–1271  
   superconductors type II, 1271–1274  
 Polarization-sensitive OCT (PSOCT), 478, 479  
 Polarized Neutron TOMography (PONTO), 1266  
 Poly-crystalline diamond (PCD)  
   fatigue process in, 577–582  
   layer, stress gradients in, 576–577  
   properties, 573–576  
   tools, 572–573  
 Poly-crystalline silicon, 349  
 Polymethyl methacrylate (PMMA) grating, 529  
 Polyoxymethylene, 754  
 Porous titanium films, 204–207  
 Positron annihilation lifetime spectroscopy (PALS)  
   constant fraction discriminator, 1315  
   detectors, 1316  
   multichannel analyzer, 1315  
   principle, 1316  
   source, 1317  
   time-to-amplitude converter, 1315  
 Positron annihilation spectroscopy (PAS)  
   DBS, 1318–1323  
   history, 1304–1305  
   lattice crystal defect, 1303  
   in molecular systems, 1311–1312  
   PLAS (*see* Positron annihilation lifetime spectroscopy (PALS))  
   schematic representation, 1306, 1307  
   sources, 1313  
   trapping model, 1308–1311  
   trapping rate and lifetime, 1306  
   vacancy-type defects, 1305  
   VEPAS (*see* Variable-energy positron annihilation spectroscopy (VEPAS))  
 Postprocessing support tools (PSTs)  
   discriminant analysis, 1019  
   feature selection, 1019–1020  
   Markov models, 1020  
   PCA, 1018  
   signal and image processing techniques, 1013–1016  
 Preisach-Mayergoyz (PM) framework, 299–301  
 Pressure grazing incidence, 25, 27, 35  
 Pressure wave (P-wave) excitation, 10  
   critical angle for, 24–25

- solution for, 24  
wave angles for, 23
- Principal component analysis (PCA), 719, 1018–1019
- Printed electronics, 496, 498
- Probability-of-detection (POD) analysis of flaws, 838
- Process monitoring, OCT, 492  
LEDs, solar cells, and displays, 495, 496  
material machining and processing, 493, 494
- Projection-slice theorem, 1232–1236
- Prony's method, 1015
- Propagation-based phase contrast imaging (PB PCI)  
batteries and fuel cells, 1068  
fracture and failure, 1066–1067  
fundamentals, 1061–1066  
geosciences, 1069  
high-speed studies of dynamic processes, 1070  
low-density materials and low-contrast boundaries, 1067–1068  
medical scaffolds and tissue engineering, 1069  
paleontology and cultural heritage, 1068–1069  
soft matter and food science, 1068
- Propagation-based phase-contrast tomography (PB PCT), 1067, 1069, 1071
- Propagation-based phase-contrast XCT, 1151–1154
- Protective coatings, 491
- Pseudo surface wave (PSAW), 335
- Pulse-compression radar, 1001
- Pulsed eddy current, 829  
imaging method, 772  
thermography (*see* Induction thermography)
- Pulsed eddy current testing (PECT), 649, 661
- Pulsed excitation, 1504
- Pulsed laser, 173
- Pulse-echo reflection microscopy, 77
- Pulse-phase transform (PPT), 1507
- Pulse thermography, 1408–1412
- Pupil plane spatial filtering technique, 519
- Q**
- Quantitative non-destructive evaluation (QNDE), 647, 648
- Quantitative XCT  
characteristics of fiber-reinforced polymers, 1138–1140
- data extraction, 1133  
data processing and visualization, 1130–1133
- fiber-reinforced polymer porosity, 1136–1138
- material simulation, 1143–1146
- metal porosity, 1134–1136
- metrological values, 1140–1141
- pore characteristics in foam, 1134
- Quantum mechanical approach, 544–546
- Quantum well infrared photo detector (QWIPs), 1392
- Quasi-static approach (QSA), 791, 796–798
- R**
- Radar interferometry (RI), 1556
- Radial analyzer, 524
- Radiation laws, 1376
- Radiometry, 929–959
- Radon transform, 1230
- Rail surface defects, 1516–1517
- Rail wheel defects, 870–872
- Raman-based distributed sensor systems, 613
- Raman effect, 542–558
- Raman line  
effect of stress on, 552  
effect of temperature on, 557–558  
width of, 552
- Raman scattering, 328, 542–591
- Raman spectroscopy  
advantages and disadvantages, 563–564  
applications, 564–591  
in corrosion processes, 585–588
- Rapid Mueller matrix polarimetry, 537–538
- Rayleigh distance, 83
- Rayleigh effect, 544
- Rayleigh scattering, 612
- Rebar corrosion, 908
- Reconstruction artifacts, 1106
- Reflection test, 940–942
- Reflectivity profiles analysis  
Downhill Simplex algorithm, 1203–1204  
film thickness determination, 1205  
Fourier approach, 1202  
magnetic thin film materials, 1205
- Region growing methods, 1132
- Relative humidity (RH), 607

- Remote field eddy current (RFEC) testing technology, 649
- Resistance spot welding
  - ultrasonic determinaion of weld nugget diameter, 90–97
  - ultrasonic determinaion of weld nugget thickness, 95–99
- Resolved shear stress (RSS), 570
- Resonant microwave microscopy probes, 947
- Resonant testing methods, 950–957
- Responsivity, 1383
- Reverse osmosis(RO), 1337–1340
- Reynolds number, 787, 798
- Richards-Wolf model, 519, 529
- Riemann space, 708
- Rose criterion, 1441
- Rotational wave equation, 9
- RSWA system of Tessonics Inc. Canada, 90
- 
- S**
- SBS, *see* Surface Brillouin scattering (SBS)
- Scalar potential, 8, 28
- Scanning acoustic microscopy
  - annular phased array, 108–111
  - 3D imaging, 101
  - focusing techniques, 101–103
  - for NDE application, 84–89
  - operating frequencies, 76
  - resistance spot weldings (*see* Resistance spot welding)
  - signal preprocessing and ultrasonic reflection tomography, 103–108
  - test specimen and X-ray CT iamge, 103
  - ultrasonic wave propagation, 78–84
- Scanning Fabry-Perot filter (SFPF), 604
- Scattered radiation, 1105
- Scattering tests, 943–945
- Search coil sensors, 830
- Self induction probe, 677
- Self-magnetic leakage fields (SMLF), 868
- Semi-analytical finite element (SAFE) method, 162–163
- Semi-opaque solids, surface Brillouin scattering in, 339
- Sensitivity, 1382
  - phase measurement, 406
  - shearing amount, 406–407
  - shearing direction, 407–411
- Sensors
  - characteristics, 1382–1383
  - photon sensors, 1388–1393
  - thermal sensors, 1383–1388
- Shear grazing incidence, 25, 27, 35
- Shear-horizontal (SH) wave solution, 14–15
- Shear waves, 11
- Shift of resonance frequency, 255–256
- Shim method, 1248
- Shim ratio, 1249
- Signal processing, NDT, 1527, 1529
  - backscattered ultrasound signal, 1528
  - classification methods, 1539
  - discrete time signal/time series, 1527–1528
  - feature extraction methods, 1534
  - feature reduction methods, 1538
  - intelligent signal processing, 1529, 1530
  - preprocessing, 1532
- Signal-to-noise ratio (SNR), 1014, 1364, 1366
- Sign index, 764
- Silicon wafer sawing and polishing, 494
- Simultaneous algebraic reconstruction technique (SART), 1258
- Simultaneous iterative reconstruction technique (SIRT), 1258
- Single-core single-winding fluxgate sensors, 858–859
- Single crystal diamond, impressions in, 566
- Single flipper method, 1248
- Single Mode Nonlinear Resonance Acoustic/Ultrasound Spectroscopy (SiMoNRAS/SiMoNRUS), 303
- Single-walled carbon nanotubes (SWCNTs), 583
- Singular value decomposition, 1535, 1568
- Skin depth effect, ECT
  - eddy current penetration, standard depth of, 657
  - excitation field strength, increase of, 661
  - increasing sensor sensitivity, at lower frequencies, 662
  - material properties, 662
  - transmitter and receiver technique, 661
- Sliding mesh technique (SMT), 792
- Slow dynamics (SD), 256, 304
- Small angle scattering, 1282–1288
- Smekal–Raman effect, 543
- Snell's law for ultrasonics, 19
- Soft *vs.* hard computing, 1531
- Solar cells, 495, 496
- Solid-liquid (SL) interface
  - bulk waves at, 4–35
  - coherence condition, 37
  - displacements and stresses at, 36
  - interface conditions, 37, 38
  - P-wave excitation, 39
  - SV-wave excitation, 39
- Solid-solid interface
  - coherence condition, 19

- interface conditions, 18–19  
potentials, displacements and stresses, 17–18
- Sonic infrared imaging for NDE, 1479–1495  
applications, 1483–1486  
defect detection in concrete structures, 1487–1488  
experimental setup, 1482–1483  
fast, wide-area inspection of composite structures, 1486  
non-linearity in, 1492–1494  
quantitative evaluation of defect depth by, 1488–1492
- Sonic IR crack detection, in civil structures, 1487
- Spatial carrier phase shifting (SCPS), 368
- Spatial coherence, 427–428
- Spatial phase shift digital shearography (SPS-DS)  
carrier-frequency, 388, 401–403  
dynamic tests, 385  
multi-channel, 398–401
- Spectral domain OCT (SD OCT), 473, 474
- Spectral interferometry, *see* Fourier domain OCT (FD OCT)
- Spectral radar, *see* Fourier domain OCT (FD OCT)
- Spectral width, 427
- Specular scatter  
ideal sample, 1184–1190  
multiple films metrology, 1201–1202  
rough or near-surface graded material, 1190–1192  
thin single films metrology, 1197–1201
- Spot welding, 1484
- SPS-DS, *see* Spatial phase shift digital shearography (SPS-DS)
- SSD, *see* Subsurface damage (SSD)
- Stefan-Boltzmann Law, 1376
- Stepped frequency continuous wave (SFCW), 1002
- Stereophotogrammetry system, 1550, 1551
- Stimulated Brillouin scattering (SBS), 614
- Straight-crested waves, 176
- Strain gauge, 742
- Strain measurement, 407
- Stress corrosion crack (SCC), 712, 718  
forward procedures, 714  
inversion procedures, 717
- Stressing  
dynamic, 412  
thermal, 412  
vacuum, 411
- Stress state evaluation, 236  
one axial stress state, 238–240  
three axial stress state, 246–247  
two axial stress state, 240–245
- Structural health monitoring (SHM), 45, 647, 1546  
definition, 1592  
offshore structures, 1600–1603  
pipes and vessels, 1593–1597  
principle of, 1592  
sensor networks, 1603–1605
- Structural integrity monitoring (SIM), 150
- Subharmonic phased array for crack evaluation (SPACE), 257
- Subharmonic phased array technique, 256–257
- Subsurface damage (SSD), 213–219
- Subsurface inspection, 849–853
- Superalloy CMSX-4, 351
- Super alloy turbine blades, sonic IR detection of fatigue cracks on, 1485
- Superconductivity quantum interference device (SQUID), 676, 679, 838–839
- Super-hard tetrahedral amorphous carbon (ta-C), 201–204
- Superlattices, 354
- Supersonic surface wave (SSAW), 335
- Supervised classification, 1539
- Surface acoustic waves, 173, 175–184  
laser generation of, 184–188  
measurement, 188–193
- Surface Brillouin scattering (SBS), 328–357  
high temperature, 343  
measurements, 341–343  
NDE applications of, 348–357  
in semi-opaque solids, 339–340
- Surface Brillouin scattering spectrum  
determination of elastic moduli  
of isotropic solids from, 344–345  
of single crystals from, 345–347  
of thin supported films from, 346–349  
extraction of material properties from, 344–349
- Surface elastodynamic Green's function, 329
- Surface-enhanced Raman spectroscopy (SERS), 590
- Surface Green's function, 334–337  
for coated solid, 335–339
- Surface heating process, 1503
- Surface phononic crystals (SPhC), 352–353
- Surface plasmons (SPs), 590
- Surface ripple-mediated Brillouin scattering, from opaque solids, 331–334
- Surface ripple scattering intensity, 334
- Surface topography, 95, 442
- Surface wave dispersion, effect of film and substrate on, 194–200

- SV-wave excitation  
     critical angle for, 26–28  
     solution for, 26  
     wave angles for, 25
- Sweeping technology, 743
- Synthetic aperture focusing techniques  
     (SAFT), 108
- Synthetic aperture radar (SAR), 933, 958, 1001
- T**
- Talbot–Lau grating interferometry (TLGI)., 1155
- Tangential magnetic field, upper  
     harmonics in, 884
- Temporal coherence, 425–427
- Temporal phase shift digital shearography  
     (TPS-DS), 385  
     4+1 phase shift algorithm, 396–398  
     4+4 phase shift algorithm, 393  
     3+3 temporal phase shift algorithm, 398  
     N+2 temporal phase shift algorithm, 398
- Temporal polarization phase shifting, 436
- Terahertz (THz) radiation, 978  
     FMCW systems, 971–973  
     imaging, 973–976  
     inline inspection system, 982  
     vs. IR, 969  
     vs. microwaves, 969  
     near-field, 979–980  
     non-destructive testing of lightweight designs, 981  
     TDS, 969–971  
     vs. ultrasonic, 969  
     vs. X-ray, 969
- Theory of back-projection, 1236–1243
- Thermal imaging cameras, 1394–1397
- Thermal nondestructive evaluation (TNDE), 1350, 1351
- Thermal penetration depth, 1500
- Thermal sensors  
     bolometers, 1386  
     heat capacity, 1384  
     maximum detectivity, 1385  
     operating principle of, 1384  
     and photon sensors, 1393  
     pyroelectric sensors, 1386  
     radiation thermocouples, 1385, 1386  
     thermal infrared imaging sensors, 1387, 1388  
     thermal infrared sensor, 1385
- Thermal-sprayed coatings, 207–213
- Thermal square wave method (TSWM), 1447–1448
- Thermal step method (TSM), 1447
- Thermal stressing, 412
- Thermal wave techniques  
     alternative thermal wave method, 1444  
     boundary conditions, 1430–1431  
     frequency response analysis, 1428–1429  
     harmonically heated slab, 1437–1438  
     instrumentation, 1448–1450  
     integrated castAl/PI/PZT piezomodules, 1461
- laser intensity modulation method  
     (*see* Laser intensity modulation method (LIMM))
- LTCC/PZT sensor-actuator modules  
     (*see* LTCC/PZT sensor-actuator modules)
- lumped model, 1432
- macro-fiber composite actuators, 1455
- multilayer structures, 1431–1432
- periodic surface temperature, 1435–1436
- piezoelectric sensors, 1470
- pyroelectric coefficient, 1422–1424
- pyroelectric current, 1424–1427
- pyroelectric current spectrum, 1427
- radiation loss, 1429
- square wave method, 1447
- step method, 1447
- temperature oscillations, 1420–1422
- thermal pulse, 1433–1434
- thermal pulse measurements, 1464
- thermal pulse methods, 1444–1447
- thermal steps, 1435
- thermoplastics-compatible piezoceramic modules, 1457
- Thermography  
     benefits, 1372  
     electromagnetic radiation spectrum, 1373  
     infrared spectral range, 1372  
     spectral ranges, 1373
- Therm-O-Matic system, 1498
- Thermoplastics-compatible Piezoceramic Modules (TPM), 1457
- Thin-film coating applications  
     coatings/paints, 491  
     pearls, 490–491  
     protective coatings/corrosion inspection, 491–492  
     tablets, 487–490
- Thin films, 583–584
- Three axial stress state evaluation, 246
- Three-dimensional digital image correlation (3D-DIC), 1547
- Three-dimensional photogrammetry, 1548
- Three-dimensional point tracking (3DPT), 1547

- Threshold-based segmentation  
  techniques, 1131
- Time-average interference microscopy,  
  370–374
- Time division multiplexing (TDM), 605
- Time domain OCT (TD OCT), 471
- Time-domain spectroscopy (TDS), 968,  
  969, 976
- Time-of-flight, 238, 247
- Time series, 1527  
  deterministic series, 1528  
  random/stochastic series, 1528
- Tip-enhanced Raman spectroscopy (TERS),  
  543, 564, 591
- Titanium nitride, on high-speed steel, 354
- T-joint structure, defect detection on, 1485
- Tomography approaches, 148
- Total external reflection, 1182
- Total internal reflection, 25, 27, 35
- Tracking-based phase-contrast  
  techniques, 1077
- Transmission line method, 949
- Transmission tests, 942–943
- Transparent media, elasto-optic scattering  
  in, 329
- Transparent solids Brillouin scattering, 329
- Transport of intensity equation (TIE), 1064,  
  1066
- Tunnel magneto-resistance (TMR) sensor,  
  682, 837  
  with two orthogonal excitation coils,  
    863–865
- Turbine blades, sonic IR detection of fatigue  
  cracks on, 1485
- Two axial stress state evaluation, 241
- Two-beam interferometry, 363–365
- Twyman-Green interferometer, 371
- Tyndall scattering, 544
- U**
- Ultrahigh frequency (UHF) method, 621
- Ultrasonic stress analysis, 238
- Ultrasonic stress evaluation, 247–248  
  clamp load of bolts, 248  
  railroad wheels, 249  
  *See also* Stress state evaluation
- Ultrasonic testing, 236
- Ultrasonic wavefield imaging (UWI)  
  acquiring and visualizing, 45–48  
  bonded plate guided waves, 61–65  
  bulk waves (*see* Aluminum plate bulk  
    waves)  
  composite panel guided waves, 57–61  
  direct arrival imaging, 51–53
- energy imaging, 50–51  
filtering frequency-wavenumber, 56
- frequency-wavenumber analysis  
  (*see* Frequency-wavenumber  
    analysis)
- phase and group velocity, 48
- structural health monitoring, 45
- ultrasonic testing, 72
- wavefield methods, 70–71
- Ultra wide band (UWB) radar, 957–959, 1001
- Underwater welded seams, 1600
- Uniaxial stress, 555–556
- Unit influence line (UIL), 1553
- Un-supervised classification/clustering, 1539
- UWI, *see* Ultrasonic wave imaging (UWI)
- V**
- Vacancy mobility and clustering, 1327–1329
- Vacuum stressing, 411
- Vanadium carbides, 350–351
- Variable-energy positron annihilation  
  spectroscopy (VEPAS)  
  advantage of, 1323
- CDB measurements, 1326
- embedded copper nanoprecipitates, 1329
- polymers, 1336–1340
- radioisotope-based positrons, 1323
- semiconductors, 1333–1336
- vacancy mobility and clustering, 1327
- Variable ranking methods, 1019
- Vector potential, 8
- Velocity-induced eddy current testing  
  (VIECT), 783
- Vibration testing, 273, 370
- Vibrothermography, 1407
- Viscoelastic coating, 157
- Vision-based system, 1578
- Vision-based techniques, 1548
- Voltage sensitivity, 1383
- Voltage standing wave ratio (VSWR), 997
- W**
- Water processing, biofouling in, 481, 483
- Wave equation, 4, 13, 30
- Wavelength selector, 562
- Wave vectors, 181
- Weak reaction approach, 797
- Welded steel power facility, detect disbonds and  
  cracks in, 1484–1485
- Welding, 1484
- Wheel surface defects, 1517–1518
- White light interferometry (WLI)

- White light interferometry (WLI) (*cont.*)  
 as auxiliary signal to measurement with monochromatic illumination, 452  
 benefits, 430  
 characteristics of interference objectives, 440  
 fringe analysis, 431–436  
 harmonically moving object, 449  
 large complex samples, 443  
 large cylinders, 444  
 large flat samples, 443–444  
 Linnik interferometer, 439–440  
 Michelson interferometer, 437  
 Mirau interferometer, 437  
 object imaging, 451  
 objectives setup for fringes, 440  
 objects under protective glass or in liquid media, 445–446  
 polarization, 442–443  
 roughness measurement, 452–455  
 shape and waviness measurement, 455  
 spectrally resolved, 456–458  
 surface features, 449–450  
 thick film and transparent plates, 448–449  
 thin film, 448  
 uses, 422
- Wien's Displacement Law, 1375
- Wire rope damages  
 parallel magnetized magnet segments and Hall sensors, 859  
 single-core single-winding fluxgate sensors, 858–859
- Wrapper methods, 1019
- X**
- X-ray computed tomography (XCT), 1148–1150  
 artifact correction method, 1108–1110  
 artifacts, 1104–1107  
 computed laminography, 1113–1114  
 cone beam geometry, 1099–1102  
 4D/in-situ, 1146–1150  
 development, 1097  
 dual energy/multi energy, 1115  
 fan beam geometry, 1102–1104  
 helical, 1113  
 industrial applications, 1117  
 in-line/high-speed, 1114  
 Lambert-Beer's law, 1098  
 limited-angle, 1113  
 metal characterization, 1121–1123  
 metrology, 1123–1129
- multi-scale tomography, 1111  
 nano-cone beam and nano, 1111–1112  
 parallel-beam synchrotron geometry, 1103  
 polymeric materials and wood, 1118–1119  
 propagation-based phase-contrast, 1151  
 quantitative, 1130–1146  
 and robot, 1114  
 simulation, 1116  
 Talbot Lau interferometer, 1155–1160
- X-ray image processing  
 anaglyph 3D images, 1050  
 averaging, 1033  
 background removal, 1032  
 computer vision problems, 1045  
 contrast enhancement, 1030  
 curve propagation, 1044  
 dual energy imaging, 1047  
 edge detection, 1040–1044  
 edge detection techniques, 1044  
 geometric transformations, 1036  
 graph partitioning methods, 1044  
 image filtering, 1038  
 image histogram, 1029  
 image stitching, 1037  
 lookup table (LUT), 1029  
 median filter, 1039  
 region-growing methods, 1044  
 segmentation, 1042–1046  
 sources, 1028  
 steps, 1028  
 unsharp mask, 1039
- X-ray imaging  
 attenuation of X-ray, 1169–1170  
 detection of X-ray, 1172–1173  
 generation of X-rays, 1170–1172  
 iterative reconstruction methods, 1175  
 micro-and nano-CT, 1175  
 projection method, 1173–1174  
 section imaging, 1174
- X-ray reflectivity (XRR)  
 semiconductor films, 1183  
 the principle of, 1184  
 thin-film technologies, 1182  
 total external reflection, 1182
- Y**
- Yoke-type magnetizer, 871  
 Young's modulus, 174, 175
- Z**
- z-invariant waves, 12