

Springer

Hand

Metro

and Te

**Springer Handbook
of Metrology and Testing**

Springer Handbook provides a concise compilation of approved key information on methods of research, general principles, and functional relationships in physics and engineering. The world's leading experts in the fields of physics and engineering will be assigned by one or several renowned editors to write the chapters comprising each volume. The content is selected by these experts from Springer sources (books, journals, online content) and other systematic and approved recent publications of physical and technical information.

The volumes will be designed to be useful as readable desk reference book to give a fast and comprehensive overview and easy retrieval of essential reliable key information, including tables, graphs, and bibliographies. References to extensive sources are provided.

Springer Handbook of Metrology and Testing

Horst Czichos, Tetsuya Saito, Leslie Smith (Eds.)

2nd edition
1017 Figures and 177 Tables



Springer

Editors

Horst Czichos
University of Applied Sciences Berlin
Germany

Tetsuya Saito
National Institute for Materials Science (NIMS)
Tsukuba, Ibaraki
Japan

Leslie Smith
National Institute of Standards and Technology (NIST)
Gaithersburg, MD
USA

ISBN: 978-3-642-16640-2 e-ISBN: 978-3-642-16641-9
DOI 10.1007/978-3-642-16641-9
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011930319

© Springer-Verlag Berlin Heidelberg 2011

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, 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.

Production and typesetting: le-tex publishing services GmbH, Leipzig
Senior Manager Springer Handbook: Dr. W. Skolaut, Heidelberg
Typography and layout: schreiberVIS, Seeheim
Illustrations: le-tex publishing services GmbH, Leipzig & Hippmann GbR, Schwarzenbruck
Cover design: eStudio Calamar Steinen, Barcelona
Cover production: WMXDesign GmbH, Heidelberg
Printing and binding: Stürtz GmbH, Würzburg

Printed on acid free paper

Springer is part of Springer Science+Business Media (www.springer.com)

61/3180/YL 5 4 3 2 1 0

Preface to the 2nd Edition

The ability to measure and to compare measurements between laboratories is one of the cornerstones of the scientific method. Globalization of research, development and manufacture has produced greatly increased attention to international standards of measurement. It is no longer sufficient to achieve internal consistency in measurements within a local laboratory or manufacturing facility; measurements must now be able to be reproduced accurately anywhere in the world.

These demands are especially intense in materials science and technology, where many characterization methods are needed during the various stages of materials and product cycles. In order for new materials to be used and incorporated into practical technology, their most important characteristics must be known well enough to justify large research and development costs. The useful properties of materials are generally responses to external fields or loads under specific conditions. The stimulus field and environmental conditions must be completely specified in order to develop a reproducible response, and to obtain reliable characteristics and data. Standard test and calibration methods describe these conditions and the Springer Handbook of Materials Measurement Methods was developed to assist scientists and engineers in both industry and academe in this task.

In this second edition of the handbook, we have responded to reader's requests for a more complete treatment of the internationally recognized formal metrology system. The book title has been changed to reflect this emphasis and the handbook organized in five parts: (A) Fundamentals of Metrology and Testing, (B) Chemical and Microstructural Analysis, (C) Materials Properties Measurement, (D) Materials Perfor-

mance Testing, (E) Modeling and Simulation Methods. The initial chapters are new and present, inter alia:

- Methodologies of measurement and testing, conformity assessment and accreditation
- Metrology principles and organization
- Quality in measurement and testing, including measurement uncertainty and accuracy.

All the remaining chapters have been brought up to date by the same distinguished international experts that produced the first edition.

The editors wish again to acknowledge the critical support and constant encouragement of the Publisher. In particular, Dr. Hubertus von Riedesel encouraged us greatly with the original concept and Dr. Werner Skolaut has done the technical editing to the highest standards of professional excellence. Finally, throughout the entire development of the handbook we were greatly aided by the able administrative support of Ms Daniela Tied.

May 2011
 Horst Czichos
 Tetsuya Saito
 Leslie Smith

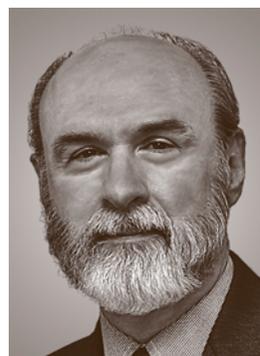
Berlin
 Tsukuba
 Washington



Horst Czichos



Tetsuya Saito



Leslie Smith

Preface to the 1st Edition

The ability to compare measurements between laboratories is one of the cornerstones of the scientific method. All scientists and engineers are trained to make accurate measurements and a comprehensive volume that provides detailed advice and leading references on measurements to scientific and engineering professionals and students is always a worthwhile addition to the literature. The principal motivation for this Springer Handbook of Materials Measurement Methods, however, stems from the increasing demands of technology for measurement results that can be used reliably anywhere in the world. These demands are especially intense in materials science and technology, where many characterization methods are needed, from scientific *composition–structure–property* relations to technological *performance–quality–reliability* assessment data, during the various stages of materials and product cycles.

In order for new materials to be used and incorporated into practical technology, their most important characteristics must be known well enough to justify large research and development costs. Furthermore, the research may be performed in one country while the engineering design is done in another and the prototype manufacture in yet another region of the world. This great emphasis on international comparison means that increasing attention must be paid to internationally recognized standards and calibration methods that go beyond careful, internally consistent, methods. This handbook was developed to assist scientists and engineers in both industry and academe in this task.

The useful properties of materials are generally responses to external fields under specific conditions.

The stimulus field and environmental conditions must be completely specified in order to develop a reproducible response. Standard test methods describe these conditions and the Chapters and an Appendix in this book contain references to the relevant international standards.

We sought out experts from all over the world that have been involved with concerns such as these. We were extremely fortunate to find a distinguished set of authors who met the challenge: to write brief chapters that nonetheless contain specific useful recommendations and resources for further information. This is the hallmark of a successful handbook. While the diverse nature of the topics covered has led to different styles of presentation, there is a commonality of purpose evident in the chapters that come from the authors' understanding of the issues facing researchers today.

This handbook would not have been possible without the visionary support of Dr. Hubertus von Riedesel, who embraced the concept and encouraged us to pursue it. We must also acknowledge the constant support of Dr. Werner Skolaut, whose technical editing has met every expectation of professional excellence. Finally, throughout the entire development of the handbook we were greatly aided by the able administrative support of Ms. Daniela Bleienberger.

March 2006
Horst Czichos
Tetsuya Saito
Leslie Smith

Berlin
Tsukuba
Washington

List of Authors

Shuji Aihara

Nippon Steel Corporation
Steel Research Laboratories
20-1, Shintomi Futtsu
293-8511 Chiba, Japan
e-mail: s-aihara@re.nsc.co.jp

Tsutomu Araki

Osaka University
Graduate School of Engineering Science
Machikaneyama, Toyonaka
560-8531 Osaka, Japan
e-mail: araki@me.es.osaka-u.ac.jp

Masaaki Ashida

Osaka University
Graduate School of Engineering Science
1-3 Machikaneyama-cho, Toyonaka
560-8531 Osaka, Japan
e-mail: ashida@mp.es.osaka-u.ac.jp

Peter D. Askew

IMSL, Industrial Microbiological Services Limited
Pale Lane
Hartley Wintney, Hants RG27 8DH, UK
e-mail: peter.askew@imsl-uk.com

Heinz-Gunter Bach

Heinrich-Hertz-Institut
Components/Integration Technology,
Heinrich-Hertz-Institute
Einsteinufer 37
10587 Berlin, Germany
e-mail: Heinz-Gunter.Bach@hhi.fraunhofer.de

Gun-Woong Bahng

Korea Research Institute of Standards and Science
Division of Chemical and Materials Metrology
Doryong-dong 1, POBox 102, Yuseoung
Daejeon, 305-600, South Korea
e-mail: gwbahng@kriss.re.kr

Claude Bathias

Conservatoire National des Arts et Métiers,
Laboratoire ITMA
Institute of Technology and Advanced Materials
2 rue Conté
75003 Paris, France
e-mail: bathias@cnam.fr

Günther Bayreuther

University of Regensburg
Physics Department
Universitätsstr. 31
93040 Regensburg, Germany
e-mail:
guenther.bayreuther@physik.uni-regensburg.de

Bernd Bertsche

University of Stuttgart
Institute of Machine Components
Pfaffenwaldring 9
70569 Stuttgart, Germany
e-mail: bernd.bertsche@ima.uni-stuttgart.de

Brian Brookman

LGC Standards Proficiency Testing
Europa Business Park, Barcroft Street
Bury, Lancashire, BL9 5BT, UK
e-mail: brian.brookman@lgcstandards.com

Wolfgang Buck

Physikalisch-Technische Bundesanstalt
Abbestrasse 2-12
10587 Berlin, Germany
e-mail: wolfgang.buck@ptb.de

Richard R. Cavanagh

National Institute of Standards
and Technology (NIST)
Surface and Microanalysis Science Division,
Chemical Science and Technology Laboratory (CSTL)
100 Bureau Drive, MS 8371
Gaithersburg, MD 20899, USA
e-mail: richard.cavanagh@nist.gov

Leonardo De Chiffre

Technical University of Denmark
Department of Mechanical Engineering
Produktionstorvet, Building 425
2800 Kgs. Lyngby, Denmark
e-mail: *ldch@mek.dtu.dk*

Steven J. Choquette

National Institute of Standards and Technology
Biochemical Science Division
100 Bureau Dr., MS 8312
Gaithersburg, MD 20899-8312, USA
e-mail: *steven.choquette@nist.gov*

Horst Czichos

University of Applied Sciences Berlin
BHT Berlin, Luxemburger Strasse 10
13353 Berlin, Germany
e-mail: *horst.czichos@t-online.de*

Werner Daum

Federal Institute for Materials Research
and Testing (BAM)
Division VIII.1 Measurement and Testing
Technology; Sensors
Unter den Eichen 87
12205 Berlin, Germany
e-mail: *werner.daum@bam.de*

Paul DeRose

National Institute of Standards and Technology
Biochemical Science Division
100 Bureau Drive, MS 8312
Gaithersburg, MD 20899-8312, USA
e-mail: *paul.derose@nist.gov*

Stephen L.R. Ellison

LGC Ltd.
Bioinformatics and Statistics
Queens Road, Teddington
Middlesex, TW11 0LY, UK
e-mail: *s.ellison@lgc.co.uk*

Anton Erhard

Federal Institute for Materials Research
and Testing (BAM)
Department Containment Systems for Dangerous
Goods
Unter den Eichen 87
12205 Berlin, Germany
e-mail: *anton.erhard@bam.de*

Uwe Ewert

Federal Institute for Materials Research
and Testing (BAM)
Division VIII.3 Radiology
Unter den Eichen 87
12205 Berlin, Germany
e-mail: *uwe.ewert@bam.de*

Richard J. Fields

National Institute of Standards and Technology
Materials Science and Engineering Laboratory
100 Bureau Drive
Gaithersburg, MD 20899, USA
e-mail: *richard.fields@nist.gov*

David Flaschenträger

Fraunhofer Institute for Structural Durability and
System Reliability (LBF)
Bartningstrasse 47
64289 Darmstadt, Germany
e-mail: *David.flaschentraeger@lbf.fraunhofer.de*

Benny D. Freeman

The University of Texas at Austin, Center for Energy
and Environmental Resources
Department of Chemical Engineering
10100 Burnet Road, Building 133, R-7100
Austin, TX 78758, USA
e-mail: *freeman@che.utexas.edu*

Holger Frenz

University of Applied Sciences Gelsenkirchen
Business Engineering
August-Schmidt-Ring 10
45665 Recklinghausen, Germany
e-mail: *frenz@fh-gelsenkirchen.de*

Jochen Gäng

JHP – consulting association for product reliability
Nobelstraße 15
70569 Stuttgart, Germany
e-mail: gaeng@jhp-beratung.de

Anja Geburtig

BAM Federal Institute for Materials Research and Testing
Division III.1, Dangerous Goods Packaging
Unter den Eichen 44-46
12203 Berlin, Germany
e-mail: anja.geburtig@bam.de

Mark Gee

National Physical Laboratory
Division of Engineering and Processing Control
Hampton Road
Teddington, TW11 0LW, UK
e-mail: mark.gee@npl.co.uk

Jürgen Goebbels

Federal Institute for Materials Research and Testing (BAM)
Radiology (VIII.3)
Unter den Eichen 87
12205 Berlin, Germany
e-mail: juergen.goebbels@bam.de

Manfred Golze

BAM Federal Institute for Materials Research and Testing
S.1 Quality in Testing
Unter den Eichen 87
12205 Berlin, Germany
e-mail: manfred.golze@bam.de

Anna A. Gorbushina

Federal Institute for Materials Research and Testing (BAM)
Department Materials and Environment
Unter den Eichen 87
12205 Berlin, Germany
e-mail: anna.gorbushina@bam.de

Robert R. Greenberg

National Institute of Standards and Technology
Analytical Chemistry Division
100 Bureau Drive, MS 8395
Gaithersburg, MD 20899-8395, USA
e-mail: robert.greenberg@nist.gov

Manfred Grinda

Landreiterweg 22
12353 Berlin, Germany
e-mail: manfred.grinda@t-online.de

Roland Grössinger

Technical University Vienna
Institut für Festkörperphysik
Wiedner Hauptstr. 8-10
Vienna, Austria
e-mail: rgroess@ifp.tuwien.ac.at

Yukito Hagihara

Sophia University
Faculty of Science and Technology, Department of Engineering and Applied Science
7-1 Kioi-cho, Chiyoda-ku
102-8554 Tokyo, Japan
e-mail: hagihara@me.sophia.ac.jp

Junhee Hahn

Korea Research Institute of Standards and Science (KRISS)
Division of Industrial Metrology
1 Doryong-dong, Yuseong-gu
Daejeon, 305-340, South Korea
e-mail: juny@kriss.re.kr

Holger Hanselka

Fraunhofer-Institute for Structural Durability and System Reliability (LBF)
Bartningstrasse 47
64289 Darmstadt, Germany
e-mail: holger.hanselka@lbf.fraunhofer.de

Werner Hässelbarth

Charlottenstr. 17A
12247 Berlin, Germany
e-mail: werner.haesselbarth@t-online.de

Martina Hedrich

BAM Federal Institute for Materials Research and Testing
S.1 Quality in Testing
Unter den Eichen 87
12205 Berlin, Germany
e-mail: martina.hedrich@bam.de

Manfred P. Hentschel

Federal Institute for Materials Research and Testing (BAM)
VIII.3, Nondestructive Testing
12200 Berlin, Germany
e-mail: manfred.hentschel@bam.de

Horst Hertel

Federal Institute for Materials Research and Testing (BAM)
Division IV.1
Unter den Eichen 87
12205 Berlin, Germany
e-mail: horst.hertel@bam.de

Daniel Hofmann

University of Stuttgart
Institute of Machine Components
Pfaffenwaldring 9
70569 Stuttgart, Germany
e-mail: daniel.hofmann@ima.uni-stuttgart.de

Xiao Hu

National Institute for Materials Science
World Premier International Center for Materials Nanoarchitectonics
Namiki 1-1
305-0044 Tsukuba, Japan
e-mail: hu.xiao@nims.go.jp

Ian Hutchings

University of Cambridge, Institute for Manufacturing
Department of Engineering
17 Charles Babbage Road
Cambridge, CB3 0FS, UK
e-mail: imh2@cam.ac.uk

Shoji Imatani

Kyoto University
Department of Energy Conversion Science
Yoshida-honmachi, Sakyo-ku
606-8501 Kyoto, Japan
e-mail: imatani@energy.kyoto-u.ac.jp

Hanspeter Ischi

The Swiss Accreditation Service (SAS)
Lindenweg 50
3003 Berne, Switzerland
e-mail: hanspeter.ischi@sas.ch

Bernd Isecke

Federal Institute for Materials Research and Testing (BAM)
Department Materials Protection and Surface Technologies
Unter den Eichen 87
12203 Berlin, Germany
e-mail: bernd.isecke@bam.de

Tadashi Itoh

Osaka University
Institute for NanoScience Design
1-3, Machikaneyama-cho, Toyonaka
560-8531 Osaka, Japan
e-mail: itoh@insd.osaka-u.ac.jp

Tetsuo Iwata

The University of Tokushima
Department of Mechanical Engineering
2-1, Minami-Jyosanjima
770-8506 Tokushima, Japan
e-mail: iwata@me.tokushima-u.ac.jp

Gerd-Rüdiger Jaenisch

Federal Institute for Materials Research and Testing (BAM)
Non-destructive Testing
Unter den Eichen 87
12205 Berlin, Germany
e-mail: gerd-ruediger.jaenisch@bam.de

Oliver Jann

Federal Institute for Materials Research
and Testing (BAM)
Environmental Material and Product Properties
Unter den Eichen 87
12205 Berlin, Germany
e-mail: oliver.jann@bam.de

Enrico Janssen

Fraunhofer Institute for Structural Durability and
System Reliability (LBF)
Bartningstrasse 47
64289 Darmstadt, Germany
e-mail: enrico.janssen@lbf.fraunhofer.de

Masanori Kohno

National Institute for Materials Science
Computational Materials Science Center
1-2-1 Sengen
305-0047 Tsukuba, Japan
e-mail: kohno.masanori@nims.go.jp

Toshiyuki Koyama

Nagoya Institute of Technology
Department of Materials Science and Engineering
Gokiso-cho, Showa-ku
466-8555 Nagoya, Japan
e-mail: koyama.toshiyuki@nitech.ac.jp

Gary W. Kramer

National Institute of Standards and Technology
Biospectroscopy Group, Biochemical Science
Division
100 Bureau Drive
Gaithersburg, MD 20899-8312, USA
e-mail: gary.kramer@nist.gov

Wolfgang E. Krumbein

Biogema
Material Ecology
Drakestrasse 68
12205 Berlin-Lichterfelde, Germany
e-mail: wek@biogema.de

George Lamaze

National Institute of Standards and Technology
Analytical Chemistry Division
100 Bureau Dr. Stop 8395
Gaithersburg, MD 20899, USA
e-mail: george.lamaze@nist.gov

Haiqing Lin

Membrane Technology and Research, Inc.
1306 Willow Road, Suite 103
Menlo Park, CA 94025, USA
e-mail: Haiqing.lin@mtrinc.com

Richard Lindstrom

National Institute of Standards and Technology
Analytical Chemistry Division
100 Bureau Drive
Gaithersburg, MD 20899-8392, USA
e-mail: richard.lindstrom@nist.gov

Samuel Low

National Institute of Standards and Technology
Metallurgy Division, Materials Science
and Engineering Laboratory
100 Bureau Drive, Mail Stop 8553
Gaithersburg, MD 20899, USA
e-mail: samuel.low@nist.gov

Koji Maeda

The University of Tokyo
Department of Applied Physics
Hongo, Bunkyo-ku
113-8656 Tokyo, Japan
e-mail: maeda@exp.t.u-tokyo.ac.jp

Ralf Matschat

Püttbergeweg 23
12589 Berlin, Germany
e-mail: ralf.matschat43@web.de

Willie E. May

National Institute of Standards
and Technology (NIST)
Chemical Science and Technology Laboratory (CSTL)
100 Bureau Drive, MS 8300
Gaithersburg, MD 20899, USA
e-mail: willie.may@nist.gov

Takashi Miyata

Nagoya University
Department of Materials Science and Engineering
464-8603 Nagoya, Japan
e-mail: miyata@numse.nagoya-u.ac.jp

Hiroshi Mizubayashi

University of Tsukuba
Institute of Materials Science
305-8573 Tsukuba, Japan
e-mail: mizuh@ims.tsukuba.ac.jp

Bernd R. Müller

BAM Federal Institute for Materials Research and Testing
Unter den Eichen 87
12205 Berlin, Germany
e-mail: bernd.mueller@bam.de

Rolf-Joachim Müller

Gesellschaft für Biotechnologische Forschung mbH
TU-BCE
Mascheroder Weg 1
Braunschweig, 38124, Germany
e-mail: rmu@gbf.de

Kiyofumi Muro

Chiba University
Faculty of Science, Department of Physics
1-33 Yayoi-cho, Inage-ku
283-8522 Chiba, Japan
e-mail: muro@physics.s.chiba-u.ac.jp

Yoshihiko Nonomura

National Institute for Materials Science
Computational Materials Science Center
Sengen 1-2-1, Tsukuba
305-0047 Ibaraki, Japan
e-mail: nonomura.yoshihiko@nims.go.jp

Jürgen Nuffer

Fraunhofer Institute for Structural Durability and System Reliability (LBF)
Bartningstrasse 47
64289 Darmstadt, Germany
e-mail: juergen.nuffer@lbf.fraunhofer.de

Jan Obrzut

National Institute of Standards and Technology
Polymers Division
100 Bureau Dr.
Gaithersburg, MD 20899-8541, USA
e-mail: jan.obrzut@nist.gov

Hiroshi Ohtani

Kyushu Institute of Technology
Department of Materials Science and Engineering
Sensui-cho 1-1, Tobata-ku
804-8550 Kitakyushu, Japan
e-mail: ohtani@matsc.kyutech.ac.jp

Kurt Osterloh

Federal Institute for Materials Research and Testing (BAM)
Division VIII.3
Unter den Eichen 97
12205 Berlin, Germany
e-mail: kurt.osterloh@bam.de

Michael Pantke

Federal Institute for Materials Research and Testing (BAM)
Division IV.1, Materials and Environment
12205 Berlin, Germany
e-mail: michael.pantke@bam.de

Karen W. Phinney

National Institute of Standards and Technology
Analytical Chemistry Division
100 Bureau Drive, Stop 8392
Gaithersburg, MD 20899-8392, USA
e-mail: karen.phinney@nist.gov

Rüdiger (Rudy) Plarre

Federal Institute for Materials Research and Testing (BAM)
Environmental Compatibility of Materials
Unter den Eichen 87
12205 Berlin, Germany
e-mail: ruediger.plarre@bam.de

Kenneth W. Pratt

National Institute of Standards and Technology
Analytical Chemistry Division
100 Bureau Dr., Stop 8391
Gaithersburg, MD 20899-8391, USA
e-mail: kenneth.pratt@nist.gov

Michael H. Ramsey

University of Sussex
School of Life Sciences
Brighton, BN1 9QG, UK
e-mail: m.h.ramsey@sussex.ac.uk

Peter Reich

Treppendorfer Weg 33
12527 Berlin, Germany
e-mail: peterreichberlin@email.de

Gunnar Ross

Magnet-Physik Dr. Steingroever GmbH
Emil-Hoffmann-Str. 3
50996 Köln, Germany
e-mail: ross@magnet-physik.de

Steffen Rudtsch

Physikalisch-Technische Bundesanstalt (PTB)
Abbestr. 2–12
10587 Berlin, Germany
e-mail: steffen.rudtsch@ptb.de

Lane Sander

National Institute of Standards and Technology
Chemical Science and Technology Laboratory,
Analytical Chemistry Division
100 Bureau Drive, MS 8392
Gaithersburg, MD 20899–8392, USA
e-mail: lane.sander@nist.gov

Erich Santner

Bürvigstraße 43
53177 Bonn, Germany
e-mail: esantner@arcor.de

Michele Schantz

National Institute of Standards and Technology
Analytical Chemistry Division
100 Bureau Drive, Stop 8392
Gaithersburg, MD 20899, USA
e-mail: michele.schantz@nist.gov

Anita Schmidt

Federal Institute for Materials Research and
Testing (BAM)
Unter den Eichen 44–46
12203 Berlin, Germany
e-mail: anita.schmidt@bam.de

Guenter Schmitt

Iserlohn University of Applied Sciences
Laboratory for Corrosion Protection
Frauenstuhlweg 31
8644 Iserlohn, Germany
e-mail: schmitt.g@fh-swf.de

Bernd Schumacher

Physikalisch-Technische Bundesanstalt
Department 2.1 DC and Low Frequency
Bundesallee 100
38116 Braunschweig, Germany
e-mail: bernd.schumacher@ptb.de

Michael Schütze

Karl-Winnacker-Institut
DECHEMA e.V.
Theodor-Heuss-Allee 25
Frankfurt am Main, 60486, Germany
e-mail: schuetze@dechema.de

Karin Schwibbert

Federal Institute for Materials Research
and Testing (BAM)
Division IV.1 Materials and Environment
12205 Berlin, Germany
e-mail: karin.schwibbert@bam.de

John H. J. Scott

National Institute of Standards and Technology
Surface and Microanalysis Science Division
100 Bureau Drive
Gaithersburg, MD 20899–8392, USA
e-mail: johnhenry.scott@nist.gov

Martin Seah

National Physical Laboratory
Analytical Science Division
Hampton Road, Middlesex
Teddington, TW11 0LW, UK
e-mail: martin.seah@npl.co.uk

Steffen Seitz

Physikalisch-Technische Bundesanstalt (PTB)
Dept. 3.13 Metrology in Chemistry
Bundesallee 100
38116 Braunschweig, Germany
e-mail: steffen.seitz@ptb.de

Masato Shimono

National Institute for Materials Science
Computational Materials Science Center
1–2–1 Sengen
305–0047 Tsukuba, Japan
e-mail: shimono.masato@nims.go.jp

John R. Sieber

National Institute of Standards and Technology
Chemical Science and Technology Laboratory
100 Bureau Drive, Stop 8391
Gaithersburg, MD 20899, USA
e-mail: john.sieber@nist.gov

Franz-Georg Simon

Federal Institute for Materials Research
and Testing (BAM)
Waste Treatment and Remedial Engineering
Unter den Eichen 87
12205 Berlin, Germany
e-mail: franz-georg.simon@bam.de

John Small

National Institute of Standards and Technology
Surface and Microanalysis Science Division
100 Bureau Drive, MS 8370
Gaithersburg, MD 20899, USA
e-mail: john.small@nist.gov

Melody V. Smith

National Institute of Standards and Technology
Biospectroscopy Group
100 Bureau Drive
Gaithersburg, MD 20899-8312, USA
e-mail: melody.smith@nist.gov

Petra Spitzer

Physikalisch-Technische Bundesanstalt (PTB)
Dept. 3.13 Metrology in Chemistry
Bundesallee 100
38116 Braunschweig, Germany
e-mail: petra.spitzer@ptb.de

Thomas Steiger

Federal Institute for Materials Research and
Testing (BAM)
Department of Analytical Chemistry; Reference
Materials
Richard-Willstätter-Straße 11
12489 Berlin, Germany
e-mail: thomas.steiger@bam.de

Ina Stephan

Federal Institute for Materials Research
and Testing (BAM)
Division IV.1 Biology in Materials Protection
and Environmental Issues
Unter den Eichen 87
12205 Berlin, Germany
e-mail: ina.stephan@bam.de

Stephan J. Stranick

National Institute of Standards and Technology
Department of Commerce, Surface
and Microanalysis Science Division
100 Bureau Dr.
Gaithersburg, MD 20899-8372, USA
e-mail: stranick@nist.gov

Hans-Henning Strehblow

Heinrich-Heine-Universität
Institute of Physical Chemistry
Universitätsstr. 1
40225 Düsseldorf, Germany
e-mail: henning@uni-duesseldorf.de

Tetsuya Tagawa

Nagoya University
Department of Materials Science and Engineering
464-8603 Nagoya, Japan
e-mail: tagawa@numse.nagoya-u.ac.jp

Akira Tezuka

National Institute of Advanced Industrial Science
and Technology
AIST Tsukuba Central 2
305-8568 Tsukuba, Japan
e-mail: tezuka.akira@aist.go.jp

Yo Tomota

Ibaraki University
Department of Materials Science,
Faculty of Engineering
4-12-1 Nakanarusawa-cho
316-8511 Hitachi, Japan
e-mail: tomota@mx.ibaraki.ac.jp

John Travis

National Institute of Standards and Technology
Analytical Chemistry Division (Retired)
100 Bureau Drive, MS 8312
Gaithersburg, MD 20899-8312, USA
e-mail: john.travis@nist.gov

Peter Trubiroha

Schlettstadter Str. 116
14169 Berlin, Germany
e-mail: peter.trubiroha@t-online.de

Gregory C. Turk

National Institute of Standards and Technology
Analytical Chemistry Division
100 Bureau Drive, Stop 8391
Gaithersburg, MD 20899, USA
e-mail: turk@nist.gov

Thomas Vetter

National Institute of Standards
and Technology (NIST)
Analytical Chemistry Division
100 Bureau Dr. Stop 8391
Gaithersburg, MD 20899-8391, USA
e-mail: thomas.vetter@nist.gov

Volker Wachtendorf

BAM Federal Institute for Materials Research and
Testing (BAM)
Division VI.3, Durability of Polymers
Unter den Eichen 87
12205 Berlin, Germany
e-mail: volker.wachtendorf@bam.de

Andrew Wallard

Pavillon de Breteuil
Bureau International des Poids et Mesures
92312 Sèvres, France
e-mail: fjoly@bipm.org

Joachim Wecker

Siemens AG
Corporate Technology, CT T DE
91050 Erlangen, Germany
e-mail: joachim.wecker@siemens.com

Wolfhard Wegscheider

Montanuniversität Leoben
General and Analytical Chemistry
Franz-Josef-Strasse 18
8700 Leoben, Austria
e-mail: wegschei@unileoben.ac.at

Alois Wehrstedt

DIN Deutsches Institut für Normung
Normenausschuss Materialprüfung (NMP)
Burggrafenstraße 6
10787 Berlin, Germany
e-mail: aloes.wehrstedt@din.de

Michael Welch

National Institute of Standards and Technology
100 Bureau Drive, Stop 8392
Gaithersburg, MD 20899, USA
e-mail: michael.welch@nist.gov

Ulf Wickström

SP Swedish National Testing and Research Institute
Department of Fire Technology
501 15 Borås, Sweden
e-mail: ulf.wickstrom@sp.se

Sheldon M. Wiederhorn

National Institute of Standards and Technology
Materials Science and Engineering Laboratory
100 Bureau Drive
Gaithersburg, MD 20899-8500, USA
e-mail: sheldon.wiederhorn@nist.gov

Scott Wight

National Institute of Standards and Technology
Surface and Microanalysis Science Division
100 Bureau Drive, Stop 8371
Gaithersburg, MD 20899-8371, USA
e-mail: scott.wight@nist.gov

Michael Winchester

National Institute of Standards and Technology
Analytical Chemistry Division
100 Bureau Drive, Building 227, Mailstop 8391
Gaithersburg, MD 20899, USA
e-mail: mrw@nist.gov

Noboru Yamada

Matsushita Electric Industrial Co., Ltd.
Optical Media Group, Storage Media Systems
Development Center
3-1-1 Yagumo-Nakamachi
570-8501 Moriguchi, Japan
e-mail: yamada.noboru@jp.panasonic.com

Rolf Zeisler

National Institute of Standards and Technology
Analytical Chemistry Division
100 Bureau Drive, MS 8395
Gaithersburg, MD 20899-8395, USA
e-mail: rolf.zeisler@nist.gov

Uwe Zscherpel

Federal Institute for Materials Research
and Testing (BAM)
Division "NDT – Radiological Methods"
Unter den Eichen 87
12205 Berlin, Germany
e-mail: uwez@bam.de

Adolf Zschunke

Rapsweg 115
04207 Leipzig, Germany
e-mail: zschunke35@aol.com

Contents

List of Abbreviations	XXV
------------------------------------	-----

Part A Fundamentals of Metrology and Testing

1 Introduction to Metrology and Testing

<i>Horst Czichos</i>	3
1.1 Methodologies of Measurement and Testing	3
1.2 Overview of Metrology	9
1.3 Fundamentals of Materials Characterization	13
References	22

2 Metrology Principles and Organization

<i>Andrew Wallard</i>	23
2.1 The Roots and Evolution of Metrology	23
2.2 BIPM: The Birth of the Metre Convention	25
2.3 BIPM: The First 75 Years	26
2.4 Quantum Standards: A Metrological Revolution	28
2.5 Regional Metrology Organizations	29
2.6 Metrological Traceability	29
2.7 Mutual Recognition of NMI Standards: The CIPM MRA	30
2.8 Metrology in the 21st Century	32
2.9 The SI System and New Science	34
References	37

3 Quality in Measurement and Testing

<i>Michael H. Ramsey, Stephen L.R. Ellison, Horst Czichos, Werner Hässelbarth, Hanspeter Ischi, Wolfhard Wegscheider, Brian Brookman, Adolf Zschunke, Holger Frenz, Manfred Golze, Martina Hedrich, Anita Schmidt, Thomas Steiger</i>	39
3.1 Sampling	40
3.2 Traceability of Measurements	45
3.3 Statistical Evaluation of Results	50
3.4 Uncertainty and Accuracy of Measurement and Testing	68
3.5 Validation	78
3.6 Interlaboratory Comparisons and Proficiency Testing	87
3.7 Reference Materials	97
3.8 Reference Procedures	116
3.9 Laboratory Accreditation and Peer Assessment	126
3.10 International Standards and Global Trade	130
3.11 Human Aspects in a Laboratory	134
3.12 Further Reading: Books and Guides	138
References	138

Part B Chemical and Microstructural Analysis

4 Analytical Chemistry

<i>Willie E. May, Richard R. Cavanagh, Gregory C. Turk, Michael Winchester, John Travis, Melody V. Smith, Paul DeRose, Steven J. Choquette, Gary W. Kramer, John R. Sieber, Robert R. Greenberg, Richard Lindstrom, George Lamaze, Rolf Zeisler, Michele Schantz, Lane Sander, Karen W. Phinney, Michael Welch, Thomas Vetter, Kenneth W. Pratt, John H. J. Scott, John Small, Scott Wight, Stephan J. Stranick, Ralf Matschat, Peter Reich</i>		145
4.1	Bulk Chemical Characterization	145
4.2	Microanalytical Chemical Characterization	179
4.3	Inorganic Analytical Chemistry: Short Surveys of Analytical Bulk Methods	189
4.4	Compound and Molecular Specific Analysis: Short Surveys of Analytical Methods	195
4.5	National Primary Standards – An Example to Establish Metrological Traceability in Elemental Analysis	198
	References	199

5 Nanoscopic Architecture and Microstructure

<i>Koji Maeda, Hiroshi Mizubayashi</i>		205
5.1	Fundamentals	211
5.2	Crystalline and Amorphous Structure Analysis	232
5.3	Lattice Defects and Impurities Analysis	239
5.4	Molecular Architecture Analysis	258
5.5	Texture, Phase Distributions, and Finite Structures Analysis	269
	References	277

6 Surface and Interface Characterization

<i>Martin Seah, Leonardo De Chiffre</i>		281
6.1	Surface Chemical Analysis	282
6.2	Surface Topography Analysis	308
	References	326

Part C Materials Properties Measurement

7 Mechanical Properties

<i>Sheldon M. Wiederhorn, Richard J. Fields, Samuel Low, Gun-Woong Bahng, Alois Wehrstedt, Junhee Hahn, Yo Tomota, Takashi Miyata, Haiqing Lin, Benny D. Freeman, Shuji Aihara, Yukito Hagihara, Tetsuya Tagawa</i>		339
7.1	Elasticity	340
7.2	Plasticity	355

7.3	Hardness	366
7.4	Strength	388
7.5	Fracture Mechanics.....	408
7.6	Permeation and Diffusion	426
	References	442
8	Thermal Properties	
	<i>Wolfgang Buck, Steffen Rudtsch</i>	453
8.1	Thermal Conductivity and Specific Heat Capacity	454
8.2	Enthalpy of Phase Transition, Adsorption and Mixing.....	462
8.3	Thermal Expansion and Thermomechanical Analysis	469
8.4	Thermogravimetry	471
8.5	Temperature Sensors	471
	References	482
9	Electrical Properties	
	<i>Bernd Schumacher, Heinz-Gunter Bach, Petra Spitzer, Jan Obrzut, Steffen Seitz</i>	485
9.1	Electrical Materials	486
9.2	Electrical Conductivity of Metallic Materials.....	493
9.3	Electrolytic Conductivity.....	498
9.4	Semiconductors.....	507
9.5	Measurement of Dielectric Materials Properties.....	526
	References	537
10	Magnetic Properties	
	<i>Joachim Wecker, Günther Bayreuther, Gunnar Ross, Roland Grössinger</i>	541
10.1	Magnetic Materials	542
10.2	Soft and Hard Magnetic Materials: (Standard) Measurement Techniques for Properties Related to the $B(H)$ Loop	546
10.3	Magnetic Characterization in a Pulsed Field Magnetometer (PFM)	567
10.4	Properties of Magnetic Thin Films.....	579
	References	585
11	Optical Properties	
	<i>Tadashi Itoh, Tsutomu Araki, Masaaki Ashida, Tetsuo Iwata, Kiyofumi Muro, Noboru Yamada</i>	587
11.1	Fundamentals of Optical Spectroscopy.....	588
11.2	Microspectroscopy	605
11.3	Magneto-optical Measurement.....	609
11.4	Nonlinear Optics and Ultrashort Pulsed Laser Application	614
11.5	Fiber Optics	626
11.6	Evaluation Technologies for Optical Disk Memory Materials.....	641
11.7	Optical Sensing	649
	References	656

Part D Materials Performance Testing

12 Corrosion

<i>Bernd Isecke, Michael Schütze, Hans-Henning Strehblow</i>	667
12.1 Background	668
12.2 Conventional Electrochemical Test Methods	671
12.3 Novel Electrochemical Test Methods	695
12.4 Exposure and On-Site Testing	699
12.5 Corrosion Without Mechanical Loading	699
12.6 Corrosion with Mechanical Loading	705
12.7 Hydrogen-Induced Stress Corrosion Cracking	714
12.8 High-Temperature Corrosion	718
12.9 Inhibitor Testing and Monitoring of Efficiency	732
References	738

13 Friction and Wear

<i>Ian Hutchings, Mark Gee, Erich Santner</i>	743
13.1 Definitions and Units	743
13.2 Selection of Friction and Wear Tests	747
13.3 Tribological Test Methods	751
13.4 Friction Measurement	754
13.5 Quantitative Assessment of Wear	759
13.6 Characterization of Surfaces and Debris	764
References	767

14 Biogenic Impact on Materials

<i>Ina Stephan, Peter D. Askew, Anna A. Gorbushina, Manfred Grinda, Horst Hertel, Wolfgang E. Krumbein, Rolf-Joachim Müller, Michael Pantke, Rüdiger (Rudy) Plarre, Guenter Schmitt, Karin Schwibbert</i>	769
14.1 Modes of Materials – Organisms Interactions	770
14.2 Biological Testing of Wood	774
14.3 Testing of Organic Materials	789
14.4 Biological Testing of Inorganic Materials	811
14.5 Coatings and Coating Materials	826
14.6 Reference Organisms	833
References	838

15 Material–Environment Interactions

<i>Franz-Georg Simon, Oliver Jann, Ulf Wickström, Anja Geburtig, Peter Trubiroha, Volker Wachtendorf</i>	845
15.1 Materials and the Environment	845
15.2 Emissions from Materials	860
15.3 Fire Physics and Chemistry	869
References	883

16 Performance Control: Nondestructive Testing and Reliability Evaluation

Uwe Ewert, Gerd-Rüdiger Jaenisch, Kurt Osterloh, Uwe Zscherpel, Claude Bathias, Manfred P. Hentschel, Anton Erhard, Jürgen Goebbels, Holger Hanselka, Bernd R. Müller, Jürgen Nuffer, Werner Daum, David Flaschenträger, Enrico Janssen, Bernd Bertsche, Daniel Hofmann, Jochen Gäng.....

.....	887
16.1 Nondestructive Evaluation	888
16.2 Industrial Radiology	900
16.3 Computerized Tomography – Application to Organic Materials	915
16.4 Computerized Tomography – Application to Inorganic Materials	921
16.5 Computed Tomography – Application to Composites and Microstructures	927
16.6 Structural Health Monitoring – Embedded Sensors.....	932
16.7 Characterization of Reliability	949
16.A Appendix	967
References	968

Part E Modeling and Simulation Methods

17 Molecular Dynamics

<i>Masato Shimono</i>	975
17.1 Basic Idea of Molecular Dynamics.....	975
17.2 Diffusionless Transformation.....	988
17.3 Rapid Solidification	995
17.4 Diffusion.....	1006
17.5 Summary	1010
References	1010

18 Continuum Constitutive Modeling

<i>Shoji Imatani</i>	1013
18.1 Phenomenological Viscoplasticity.....	1013
18.2 Material Anisotropy	1018
18.3 Metallothermomechanical Coupling	1023
18.4 Crystal Plasticity	1026
References	1030

19 Finite Element and Finite Difference Methods

<i>Akira Tezuka</i>	1033
19.1 Discretized Numerical Schemes for FEM and FDM.....	1035
19.2 Basic Derivations in FEM and FDM	1037
19.3 The Equivalence of FEM and FDM Methods	1041
19.4 From Mechanics to Mathematics: Equilibrium Equations and Partial Differential Equations	1042
19.5 From Mathematics to Mechanics: Characteristic of Partial Differential Equations	1047

19.6	Time Integration for Unsteady Problems	1049
19.7	Multidimensional Case	1051
19.8	Treatment of the Nonlinear Case	1055
19.9	Advanced Topics in FEM and FDM	1055
19.10	Free Codes	1059
	References	1059
20	The CALPHAD Method	
	<i>Hiroshi Ohtani</i>	1061
20.1	Outline of the CALPHAD Method	1062
20.2	Incorporation of the First-principles Calculations into the CALPHAD Approach	1066
20.3	Prediction of Thermodynamic Properties of Compound Phases with First-principles Calculations	1079
	References	1090
21	Phase Field Approach	
	<i>Toshiyuki Koyama</i>	1091
21.1	Basic Concept of the Phase-Field Method	1092
21.2	Total Free Energy of Microstructure	1093
21.3	Solidification	1102
21.4	Diffusion-Controlled Phase Transformation	1105
21.5	Structural Phase Transformation	1108
21.6	Microstructure Evolution	1110
	References	1114
22	Monte Carlo Simulation	
	<i>Xiao Hu, Yoshihiko Nonomura, Masanori Kohno</i>	1117
22.1	Fundamentals of the Monte Carlo Method	1117
22.2	Improved Algorithms	1121
22.3	Quantum Monte Carlo Method	1126
22.4	Bicritical Phenomena in O(5) Model	1133
22.5	Superconductivity Vortex State	1137
22.6	Effects of Randomness in Vortex States	1143
22.7	Quantum Critical Phenomena	1146
	References	1149
	Acknowledgements	1159
	About the Authors	1161
	Detailed Contents	1185
	Subject Index	1203

CCQM	Comité Consultative pour la Quantité de Matière	CNRS	Centre National de la Recherche Scientifique
CCQM	Consultative Committee for Quantity of Matter Metrology in Chemistry	COMAR	Code d'Indexation des Matériaux de Référence
CCRI	Consultative Committee for Ionizing Radiation	COSY	correlated spectroscopy
CCT	Consultative Committee for Thermometry	CPAA	charged particle activation analysis
CCT	center-cracked tension	CPP	current-perpendicular-to-plane
cct	crevice corrosion temperature	cpt	critical pitting temperature
CCTF	Consultative Committee for Time and Frequency	CR	computed radiography
CCU	Consultative Committee for Units	CRM	certified reference material
CD	circular dichroism	CT	compact tension
CE	Communauté Européenne	CT	compact test
CE	Conformité Européenne	CT	computed tomography
CE	capillary electrophoresis	CTD	charge transfer device
CE	counter electrode	CTE	coefficient of thermal expansion
CEM	cluster expansion method	CTOD	crack-tip opening displacement
CEN	European Committee for Standardization	CTS	collaborative trial in sampling
CEN	European Standard Organization	CVD	chemical vapor deposition
CENELEC	European Electrotechnical Standardization Commission	CVM	cluster variation method
CERT	constant extension rate test	CW	continuous wave
CFD	computational fluid dynamics	CZE	capillary zone electrophoresis
CFL	Courant–Friedrichs–Lewy		
CFRP	carbon-fiber-reinforced polymer	D	
CG	coarse grained	DA	differential amplifier
CG	conjugate gradient	DA	drop amplifier
CGE	capillary gel electrophoresis	DAD	diode-array detector
CGHE	carrier gas hot extraction	DBTT	ductile-to-brittle transition temperature
CGPM	General Conference on Weights and Measures	DC	direct current
CI	carbonyl index	DCM	double crystal monochromator
CIEF	capillary isoelectric focusing	DDA	digital detector array
CIP	constrained interpolated profile	DEM	discrete element method
CIP	current-in-plane	DF	dark field
CIPM	Comité Internationale des Poids et Mesures	DFG	difference frequency generation
CIPM	International Committee of Weights and Measures	DFT	discrete Fourier transform
CITAC	Cooperation for International Traceability in Analytical Chemistry	DI	designated institute
CITP	capillary isotachopheresis	DIN	Deutsches Institut für Normung
CL	cathodoluminescence	DIR	digital industrial radiology
CLA	center line average	DLTS	deep-level transient spectroscopy
CM	ceramic matrix	DMM	double multilayer monochromator
CMA	cylindrical mirror analyzer	DMRG	density-matrix renormalization group
CMC	calibration and measurement capability	DMS	diluted magnetic semiconductor
CMC	ceramic matrix composite	DNA	deoxyribonucleic acid
CMM	coordinate-measuring machine	DNPH	2,4-dinitrophenylhydrazine
CMN	cerium magnesium nitrate	DOC	dissolved organic carbon
CMOS	complementary metal-oxide-semiconductor	DOS	density of states
CNR	contrast-to-noise ratio	DRP	dense random packing
		DS	digital storage
		DSC	differential scanning calorimeter
		DT	tomography density
		DTA	differential thermal analysis
		DTU	Danmarks Tekniske Universitet
		DWDD	domain-wall displacement detection
		DWDM	dense wavelength-division multiplexed

J			
JCGM	Joint Committee for Guides in Metrology	MECC	micellar electrokinetic capillary chromatography
JCTLM	Joint Committee for Traceability in Laboratory Medicine	MEIS	medium-energy ion scattering
JIS	Japanese Institute of Standards	MEM	maximum entropy method
		MEMS	microelectromechanical system
		MFC	mass-flow controller
		MFD	mode field diameter
		MFM	magnetoforce micrometer
		MIBK	methylisobutylketone
		MIC	microbially induced corrosion
		MID	measuring instruments directive
		MIS	metal–insulator–semiconductor
		MISFET	metal–insulator–semiconductor FET
		MITI	Ministry of International Trade and Industry
			meter, kilogram, and second
		MKS	meter, kilogram, second, and ampere
		MKSA	meter, kilogram, second, and ampere
		MLA	multilateral agreement
		MLE	maximum-likelihood estimation
		MLLSQ	multiple linear least squares
		MM	metal matrix
		MMC	metal matrix composite
		MMF	minimum mass fraction
		MO	magneto-optical
		MOE	modulus of elasticity
		MOKE	magneto-optic Kerr effect
		MOL	magneto-optical layer
		MON	monochromator
		MOS	metal–oxide–semiconductor
		MPA	Materialprüfungsamt
		MRA	multiregional agreement
		MRA	mutual recognition arrangement
		MRAM	magnetic random-access memory
		MRI	magnetic resonance imaging
		MRR	median rank regression
		MS	magnetic stirring
		MS	mass spectrometry
		MST	microsystems technology
		MSW	municipal solid waste
		MTJ	magnetic tunnel junction
		MUT	material under test
		MUVU	mobile UV unit
		MXCD	magnetic x-ray circular dichroism
		MoU	memorandum of understanding
K			
KC	key comparison		
KCDB	key comparison database		
KKS	Kim–Kim–Suzuki		
KLM	Kerr-lens mode-locking		
KT	Kosterlitz–Thouless		
L			
LBB	Ladyzhenskaya–Babuska–Brezzi		
LBZ	local brittle zone		
LC	liquid chromatography		
LC	liquid crystal		
LCF	low cycle fatigue		
LD	Lawrence–Doniach		
LD	laser device		
LD	laser diode		
LDOS	local density of states		
LDV	laser Doppler velocimeter		
LED	light-emitting diode		
LEFM	linear-elastic fracture mechanics		
LGC	Laboratory of the Government Chemist		
LIF	laser-induced fluorescence		
LJ	Lennard-Jones		
LNE	Laboratoire nationale de métrologie et d'essais		
LOC	limit of decision		
LOD	limit of detection		
LOD	limit of determination		
LOQ	limit of quantification		
LSCM	laser scanning confocal microscope		
LSDA	local spin density approximation		
LST	linear system theory		
LTS	low-temperature superconductor		
LVDT	linear variable differential transformer		
LVM	local vibrational mode		
M			
MAD	median absolute deviation		
MC	Monte Carlo		
MCA	multichannel analyzer		
MCD	magnetic circular dichroism		
MCDA	magnetic circular dichroic absorption		
MCP	microchannel plate		
MCPE	magnetic circular-polarized emission		
MD	molecular dynamics		
MDM	minimum detectable mass		
N			
		NA	numerical aperture
		NAA	neutron activation analysis
		NAB	national accreditation body
		NACE	National Association of Corrosion Engineers
		NAFTA	North America Free Trade Association
		NBS	National Bureau of Standards
		ND	neutron diffraction
		NDE	nondestructive evaluation

NDP	neutron depth profiling	PC	personal computer
NDT	nondestructive testing	PC	photoconductive detector
NEP	noise-equivalent power	PC	polycarbonate
NEXAFS	near-edge x-ray absorption fine structure	PCB	polychlorinated biphenyl
NFPA	National Fire Protection Association	PCF	photonic crystal fiber
NHE	normal hydrogen electrode	PCI	phase contrast imaging
NIR	near infrared	PCR	polymerase chain reaction
NIST	National Institute of Standards and Technology	PDMS	poly(dimethylsiloxane)
NMI	National Metrology Institute	PE	polyethylene
NMR	nuclear magnetic resonance	PE-HD	high-density polyethylene
NMi	Netherlands Measurement Institute	PE-LD	low-density polyethylene
NOE	nuclear Overhauser effect	PEELS	parallel electron energy loss spectroscopy
NPL	National Physical Laboratory	PEM	photoelectromagnetic
NPT	number pressure temperature	PERSF	pure element relative sensitivity factor
NR	natural rubber	PET	polyethylene terephthalate
NR	neutron reflectance	PFM	pulse field magnetometer
NRA	nuclear reaction analysis	PGAA	prompt gamma activation analysis
NRC-CRM	National Research Center for Certified Reference Materials	PHB	poly(β -hydroxy butyrate)
NRW	Nordrhein-Westfalen	PI	pitting index
NTC	negative temperature coefficient	PID	photoionization detector
		PIRG	path-integral renormalization group
		PIXE	particle-induced x-ray emission
		PL	photoluminescence
O		PLE	PL excitation
		PLZT	lanthanide-modified piezoceramic
OA	operational amplifier	PM	polymer matrix
OCT	optical coherence tomography	PMMA	poly(methyl methacrylate)
ODD	object-to-detector distance	PMT	photomultiplier tube
ODF	orientation distribution function	POD	probability of detection
ODMR	optically detected magnetic resonance	POF	polymer optical fiber
ODS	octadecylsilane	POL	polychromator
OES	optical emission spectroscopy/spectrometry	POM	particulate organic matter
OIML	International Organization of Legal Metrology	POS	proof-of-screen
OKE	optical Kerr effect	PSD	power-spectral density
OM	optical microscopy	PSDF	power spectral density function
OMH	Orzajos Meresugyi Hivatal	PSI	phase-shift interferometry
OPA	optical parametric amplifier	PSL	photostimulated luminescence
OPG	optical parametric generation	PT	phototube
OPO	optical parametric oscillator	PTB	Physikalisch-Technische Bundesanstalt
OR	optical rectification	PTC	positive temperature coefficient
ORD	optical rotary dispersion	PTFE	polytetrafluoroethylene
OSA	optical spectrum analyzer	PTMSP	poly(1-trimethylsilyl-1-propyne)
OSU	Ohio State University	PU	polyurethane
OTDR	optical time-domain reflectometry	PUF	polyurethane foam
		PV	photovoltaic
P		PVA	polyvinyl acetate
		PVC	polyvinyl chloride
PA	polyamide	PVD	physical vapor deposition
PAA	photon activation analysis	PVDF	polyvinylidene fluoride
PAC	Pacific Accreditation Cooperation	PWM	pulse-width modulation
PAC	perturbed angular correlation	PZT	lead zirconate titanate
PAH	polycyclic aromatic hydrocarbon		
PAS	positron annihilation spectroscopy	Q	
PBG	photonic band gap	QA	quality assurance
		QC	quality control

QE quantum effect
 QMR quasiminimal residual
 QMS quality management system
 QNMR quantitative proton nuclear magnetic resonance

R

RAPD random amplified polymorphic DNA
 RBS Rutherford backscattering
 RC resistor–capacitor
 RD rolling direction
 RDE rotating disc electrode
 RE reference electrode
 RF radiofrequency
 RFLP restriction fragment length polymorphism
 RG renormalization group
 RH relative humidity
 RI refractive index
 RM reference material
 RMO regional metrology organization
 RMR RM report
 RMS root mean square
 RNA nuclear reaction analysis
 RNAA radiochemical NAA
 RPLC reversed-phase liquid chromatography
 RRDE rotating ring-disc electrode
 rRNA ribosomal RNA
 RSF relative sensitivity factor

S

S/N signal-to-noise ratio
 SABS South African Bureau of Standards
 SAD selected area diffraction
 SADC MET Southern African Development Community Cooperation in Measurement Traceability
 SAMR small-angle magnetization-rotation
 SAQCS sampling and analytical quality control scheme
 SAXS small-angle x-ray scattering
 SBI single burning item
 SBR styrene butyl rubber
 SBS sick-building syndrome
 SC superconductivity
 SCA surface chemical analysis
 SCC stress corrosion cracking
 SCE saturated calomel electrode
 SCLM scanning confocal laser microscopy
 SD strength difference
 SDD silicon drift detector
 SE secondary electron
 SEC specific energy consumption
 SECM scanning electrochemical microscope
 SEI secondary electron imaging

SEM scanning electron microscopy
 SEN single-edge notched
 SENB4 four-point single-edge notch bend
 SER specific emission rate
 SFG sum frequency generation
 SFM scanning force microscopy
 SHE standard hydrogen electrode
 SHG second-harmonic generation
 SHM structural health monitoring
 SI International System of Units
 SI Système International d'Unités
 SIM Sistema Interamericano de Metrología
 SIMS secondary ion mass spectrometry
 SMSC study semiconductor
 SMU Slovenski Metrologicky Ustav
 SNOM scanning near-field optical microscopy
 SNR signal-to-noise ratio
 SOD source-to-object distance
 SOLAS safety of life at sea
 SOLZ second-order Laue zone
 SOP standard operating procedure
 SOR successive overrelaxation
 SP Swedish National Testing and Research Institute
 SPD singular point detection
 SPF superplastic forming
 SPH smooth particle hydrodynamics
 SPI selective polarization inversion
 SPM scanning probe microscopy
 SPM self-phase modulation
 SPOM surface potential microscope
 SPRT standard platinum resistance thermometer
 SPT sampling proficiency test
 SQUID superconducting quantum interference device
 SRE stray radiant energy
 SRET scanning reference electrode technique
 SRM standard reference material
 SRS stimulated Raman scattering
 SS spectral sensitivity
 SSE stochastic series expansion
 SST single-sheet tester
 SST system suitability test
 STEM scanning transmission electron microscopy
 STL stereolithographic data format
 STM scanning tunneling microscopy
 STP steady-state permeation
 STS scanning tunneling spectroscopy
 SUPG streamline-upwind Petrov–Galerkin
 SVET scanning vibrating electrode technique
 SVOC semi-volatile organic compound
 SW Swendsen–Wang
 SWLI scanning white-light interferometry
 SZ stretched zone
 SZW stretched zone width

T		V	
TAC	time-to-amplitude converter	VAMAS	Versailles Project on Advanced Materials and Standards
TBCCO	tellurium-barium-calcium-copper-oxide	VCSEL	vertical-cavity surface-emitting laser
TBT	technical barriers to trade	VDEh	Verein Deutscher Eisenhttenleute
TCD	thermal conductivity detector	VG	vortex glass
TCSPC	time-correlated single-photon counting	VIM	international vocabulary of basic and general terms in metrology
TDI	time-delayed integration	VIM	international vocabulary of metrology
TDS	thermal desorption mass spectrometry	VL	vortex liquid
TDS	total dissolved solid	VOC	volatile organic carbon
TEM	transmission electron microscopy	VOST	volatile organic sampling train
TFT	thin-film transistor	VSM	vibrating-sample magnetometer
TG	thermogravimetry	VVOC	very volatile organic compound
TGA-IR	thermal gravimetric analysis-infrared		
TGFSR	twisted GFSR	W	
THG	third-harmonic generation	WDM	wavelength division multiplexing
TIMS	thermal ionization mass spectrometry	WDS	wavelength-dispersive spectrometry
TIRFM	total internal reflection fluorescence microscopy	WE	working electrode
TLA	thin-layer activation	WFI	water for injection
TMA	thermomechanical analysis	WGMM	Working Group on Materials Metrology
TMR	tunnel magneto-resistance	WHO	World Health Organization
TMS	tetramethylsilane	WLI	white-light interferometry
TOF	time of flight	WTO	World Trade Organization
TPA	two-photon absorption	WZW	Wess–Zumino–Witten
TR	technical report	X	
TRIP	transformation induced plasticity	XAS	x-ray absorption spectroscopy
TS	tensile strength	XCT	x-ray computed tomography
TTT	time–temperature-transformation	XEDS	energy-dispersive x-ray spectrometry
TU	Technical University	XFL	photoemitted Fermi level
TVOC	total volatile organic compound	XMA	x-ray micro analyzer
TW	thermostat water	XMCD	x-ray magnetic circular dichroism
TWA	technical work area	XPS	x-ray photoelectron spectroscopy
TWIP	twinning induced plasticity	XPS	x-ray photoemission spectroscopy
TXIB	2,2,4-trimethyl-1,3-pentanediol diisobutyrate	XRD	x-ray diffraction
TXRF	total reflection x-ray fluorescence spectrometry	XRF	x-ray fluorescence
		XRT	x-ray topography
U		Y	
UBA	Bundesumweltamt	YAG	yttrium aluminum garnet
UHV	ultra-high vacuum	YIG	yttrium-iron garnet
UIC	Union Internationale des Chemins de Fer	YS	yield strength
ULSI	ultralarge-scale integration	Z	
USAXS	ultrasmall-angle scattering	ZOLZ	zero-order Laue zone
USP	United States Pharmacopeia	ZRA	zero-resistance ammetry
UT	ultrasonic technique		
UTS	ultimate tensile strength		
UV	ultraviolet		
UVSG	UV Spectrometry Group		
UXO	unexploded ordnance		

Part A Fundamentals

Part A Fundamentals of Metrology and Testing

1 Introduction to Metrology and Testing

Horst Czichos, Berlin, Germany

2 Metrology Principles and Organization

Andrew Wallard, Sèvres, France

3 Quality in Measurement and Testing

Michael H. Ramsey, Brighton, UK

Stephen L.R. Ellison, Middlesex, UK

Horst Czichos, Berlin, Germany

Werner Hässelbarth, Berlin, Germany

Hanspeter Ischi, Berne, Switzerland

Wolfhard Wegscheider, Leoben, Austria

Brian Brookman, Bury, Lancashire, UK

Adolf Zschunke, Leipzig, Germany

Holger Frenz, Recklinghausen, Germany

Manfred Golze, Berlin, Germany

Martina Hedrich, Berlin, Germany

Anita Schmidt, Berlin, Germany

Thomas Steiger, Berlin, Germany

1. Introduction to Metrology and Testing

This chapter reviews the methodologies of measurement and testing. It gives an overview of metrology and presents the fundamentals of materials characterization as a basis for

1. Chemical and microstructural analysis
2. Materials properties measurement
3. Materials performance testing

which are treated in parts B, C, and D of the handbook.

1.1 Methodologies of Measurement and Testing	3
1.1.1 Measurement	3
1.1.2 Testing.....	5

1.1.3 Conformity Assessment and Accreditation.....	7
1.2 Overview of Metrology	9
1.2.1 The Meter Convention	9
1.2.2 Categories of Metrology.....	9
1.2.3 Metrological Units	11
1.2.4 Measurement Standards	12
1.3 Fundamentals of Materials Characterization	13
1.3.1 Nature of Materials.....	13
1.3.2 Types of Materials	15
1.3.3 Scale of Materials	16
1.3.4 Properties of Materials	17
1.3.5 Performance of Materials	19
1.3.6 Metrology of Materials	20
References	22

In science and engineering, objects of interest have to be characterized by measurement and testing. Measurement is the process of experimentally obtaining quantity values that can reasonably be attributed to a property of

a body or substance. Metrology is the science of measurement. Testing is the technical procedure consisting of the determination of characteristics of a given object or process, in accordance with a specified method [1.1].

1.1 Methodologies of Measurement and Testing

The methodologies of measurement and testing to determine characteristics of a given object are illustrated in a unified general scheme in Fig. 1.1, which is discussed in the next sections.

1.1.1 Measurement

Measurement begins with the definition of the measurand, the quantity intended to be measured. The specification of a measurand requires knowledge of the kind of quantity and a description of the object carrying the quantity. When the measurand is defined, it must be related to a measurement standard, the realization of the definition of the quantity to be measured. The measurement procedure is a detailed description

of a measurement according to a measurement principle and to a given measurement method. It is based on a measurement model, including any calculation to obtain a measurement result. The basic features of a measurement procedure are the following [1.1].

- Measurement principle: the phenomenon serving as a basis of a measurement
- Measurement method: a generic description of a logical organization of operations used in a measurement
- Measuring system: a set of one or more measuring instruments and often other devices, including any reagent and supply, assembled and adapted to give information used to generate measured quan-

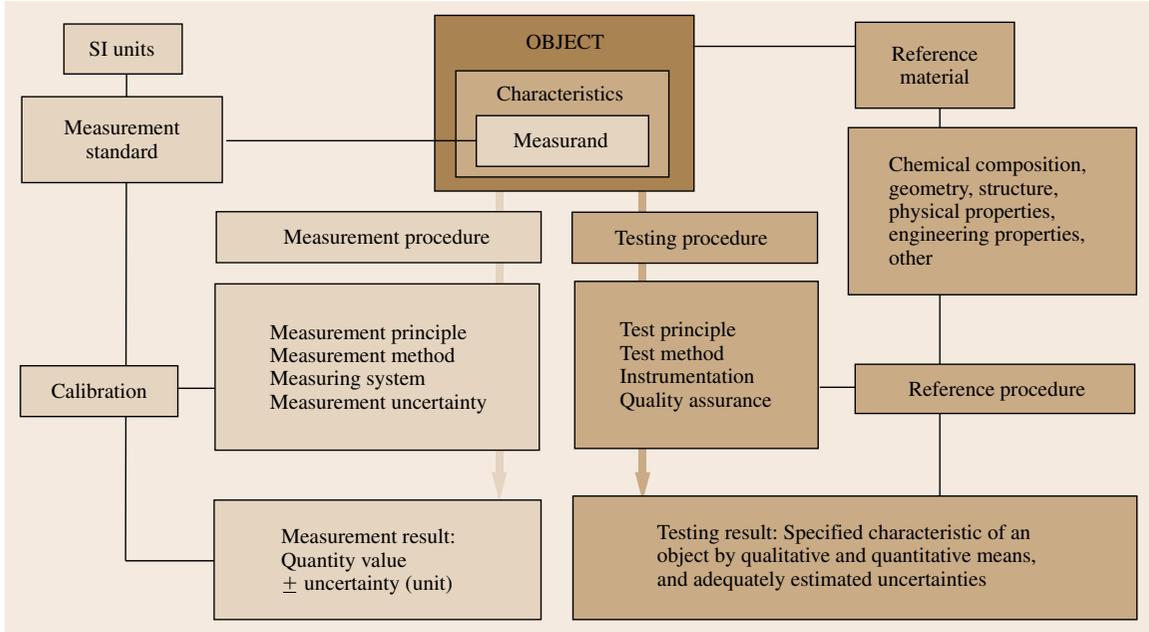


Fig. 1.1 The methodologies of measurement (light brown) and testing (dark brown) – a general scheme

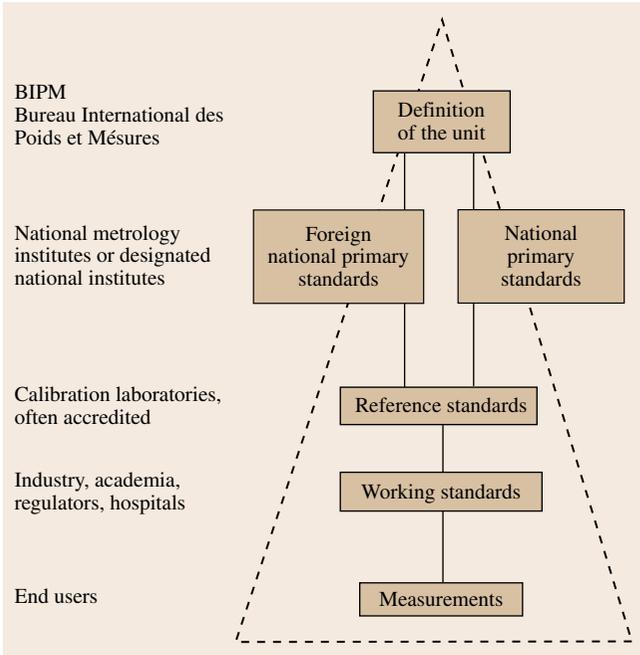


Fig. 1.2 The traceability chain for measurements

- Measurement uncertainty: a nonnegative parameter characterizing the dispersion of the quantity values being attributed to a measurand

The result of a measurement has to be expressed as a quantity value together with its uncertainty, including the unit of the measurand.

Traceability and Calibration

The measured quantity value must be related to a reference through a documented unbroken *traceability chain*. The *traceability of measurement* is described in detail in Sect. 3.2. Figure 1.2 illustrates this concept schematically.

The traceability chain ensures that a measurement result or the value of a standard is related to references at the higher levels, ending at the primary standard, based on the International System of Units (*le Système International d’Unités, SI*) (Sect. 1.2.3). An end user may obtain traceability to the highest international level either directly from a national metrology institute or from a secondary calibration laboratory, usually an accredited laboratory. As a result of various mutual recognition arrangements, internationally recognized traceability may be obtained from laboratories outside the user’s own country. Metrological timelines in traceability, defined as changes, however slight, in

tity values within specified intervals for quantities of specified kinds

all instruments and standards over time, are discussed in [1.2].

A basic tool in ensuring the traceability of a measurement is either the calibration of a measuring instrument or system, or through the use of a reference material. Calibration determines the performance characteristics of an instrument or system before its use, while reference material calibrates the instrument or system at time of use. Calibration is usually achieved by means of a direct comparison against measurement standards or certified reference materials and is documented by a calibration certificate for the instrument.

The expression “traceability to the SI” means traceability of a measured quantity value to a unit of the International System of Units. This means metrological traceability to a *dematerialized reference*, because the SI units are conceptually based on *natural constants*, e.g., the speed of light for the unit of length. So, as already mentioned and shown in Fig. 1.1, the characterization of the measurand must be realized by a *measurement standard* (Sect. 1.2.4). If a measured quantity value is an attribute of a *materialized object* (e.g., a chemical substance, a material specimen or a manufactured product), also an object-related traceability (speciation) to a *materialized reference* (Fig. 1.1) is needed to characterize the object that bears the metrologically defined and measured quantity value.

Uncertainty of Measurements

Measurement uncertainty comprises, in general, many components and can be determined in different ways [1.3]. *The Statistical Evaluation of Results* is explained in detail in Sect. 3.3, and the *Accuracy and Uncertainty of Measurement* is comprehensively described in Sect. 3.4. A basic method to determine uncertainty of measurements is the *Guide to the expression of uncertainty in measurement (GUM)* [1.4], which is shared jointly by the Joint Committee for Guides in Metrology (JCGM) member organizations (BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML). The concept of the GUM can be briefly outlined as follows [1.5].

The GUM Uncertainty Philosophy.

- A measurement quantity X , whose value is not known exactly, is considered as a stochastic variable with a probability function.
- The result x of measurement is an estimate of the expectation value $E(X)$.

- The standard uncertainty $u(x)$ is equal to the square root of an estimate of the variance $V(X)$.
- *Type A uncertainty evaluation.* Expectation and variance are estimated by statistical processing of repeated measurements.
- *Type B uncertainty evaluation.* Expectation and variance are estimated by other methods than those used for type A evaluations. The most commonly used method is to assume a probability distribution, e.g., a rectangular distribution, based on experience or other information.

The GUM Method Based on the GUM Philosophy.

- Identify all important components of measurement uncertainty. There are many sources that can contribute to measurement uncertainty. Apply a model of the actual measurement process to identify the sources. Use measurement quantities in a mathematical model.
- Calculate the standard uncertainty of each component of measurement uncertainty. Each component of measurement uncertainty is expressed in terms of the standard uncertainty determined from either a type A or type B evaluation.
- Calculate the combined uncertainty u (the *uncertainty budget*). The combined uncertainty is calculated by combining the individual uncertainty components according to the law of propagation of uncertainty. In practice
 - for a sum or a difference of components, the combined uncertainty is calculated as the square root of the sum of the squared standard uncertainties of the components;
 - for a product or a quotient of components, the same *sum/difference* rule applies as for the relative standard uncertainties of the components.
- Calculate the expanded uncertainty U by multiplying the combined uncertainty with the coverage factor k .
- State the measurement result in the form $X = x \pm U$.

The methods to determine uncertainties are presented in detail in Sect. 3.4.

1.1.2 Testing

The aim of testing is to determine characteristics (attributes) of a given object and express them by qualitative and quantitative means, including adequately

estimated uncertainties, as outlined in the right-hand side of Fig. 1.1. For the testing methodology, metrology delivers the basis for the comparability of test results, e.g., by defining the units of measurement and the associated uncertainty of the measurement results. Essential tools supporting testing include reference materials, certified reference materials, and reference procedures.

- Reference material (RM) [1.6]: a material, sufficiently homogeneous and stable with regards to specified properties, which has been established to be fit for its intended use in measurement or in examination of nominal properties
- Certified reference material (CRM): a reference material, accompanied by documentation issued by an authoritative body and providing one or more specified property values with associated uncertainties and traceabilities, using a valid procedure
- Reference procedures [1.5]: procedures of testing, measurement or analysis, thoroughly characterized and proven to be under control, intended for
 - quality assessment of other procedures for comparable tasks, or
 - characterization of reference materials including reference objects, or
 - determination of reference values.

The uncertainty of the results of a reference procedure must be adequately estimated and appropriate for the intended use. Recommendations/guides for the de-

termination of uncertainties in different areas of testing include

- Guide for the estimation of measurement uncertainty in testing [1.7]
- Guide to the evaluation of measurement uncertainty for quantitative tests results [1.8]
- Guide for chemistry [1.9]
- Measurement uncertainty in environmental laboratories [1.10]
- Uncertainties in calibration and testing [1.11].

The methodology of testing combined with measurement is exemplified in Fig. 1.3 for the determination of mechanical characteristics of a technical object.

Generally speaking, the mechanical properties of materials characterize the response of a material sample to loading. The mechanical loading action on materials in engineering applications can basically be categorized as tension, compression, bending, shear or torsion, which may be static or dynamic. In addition, thermomechanical loading effects can occur. The testing of mechanical properties consists of measuring the mechanical loading stress (force/cross-sectional area = F/A) and the corresponding materials response (strain, elongation) and expressing this as a *stress–strain curve*. Its regimes and data points characterize the mechanical behavior of materials.

Consider for example *elasticity*, which is an important characteristic of all components of engineered structures. The elastic modulus (E) describes the rela-

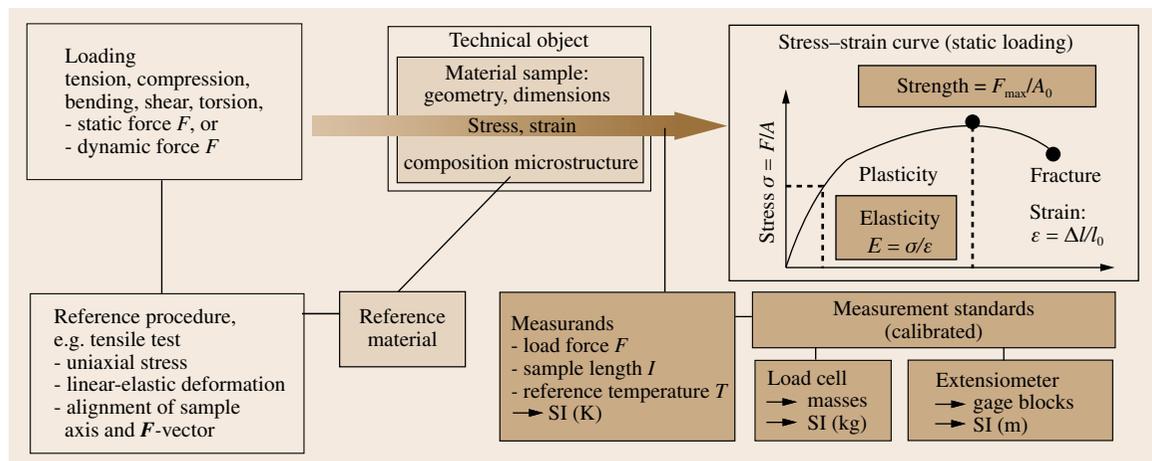


Fig. 1.3 The combination of measurement and testing to determine mechanical characteristics of the technical object

tion between a stress (σ) imposed on a material and the strain (ϵ) response of the material, or vice versa. The stimulus takes the form of an applied load, and the measured effect is the resultant displacement. The traceability of the stress is established through the use of a calibrated load cell and by measuring the specimen cross-sectional area with a calibrated micrometer, whereas the traceability of the strain is established by measuring the change in length of the originally measured gage length, usually with a calibrated strain gage. This, however, is not sufficient to ensure repeatable results unless a testing reference procedure, e.g., a standardized tensile test, is used on identically prepared specimens, backed up by a reference material.

Figure 1.3 illustrates the metrological and technological aspects.

- Metrologically, the measurands of the strength value are the force (F), area (A), and the length measurement (l) of the technical object, all at a reference temperature (T).
- Technologically and concerning testing, the mechanical characteristics expressed in a stress–strain curve depend on at least the following groups of influencing parameters, to be backed up by appropriate references.
 - The chemical and physical nature of the object: chemical composition, microstructure, and *structure–property relations* such as crystallographic shape-memory effects [1.12]; for example, strength values of metals are significantly influenced by alloying elements, grain size (fine/coarse), work-hardening treatment, etc.
 - The mechanical loading action and dependence on deformation amplitude: tension, compression, bending, shear, and torsion; for example, tensile strength is different from shear strength for a given material.
 - The time dependence of the loading mode forces (static, dynamic, impact, stochastic) and deviations from simple linear-elastic deformation (anelastic, viscoelastic or micro-viscoplastic deformation). Generally, the dynamic strength of a material is different from its static strength.

The combined measurement and testing methodologies, their operating parameters, and the traceability requirements are illustrated in a highly simplified scheme by the *confidence ring* [1.13] shown in Fig. 1.4.

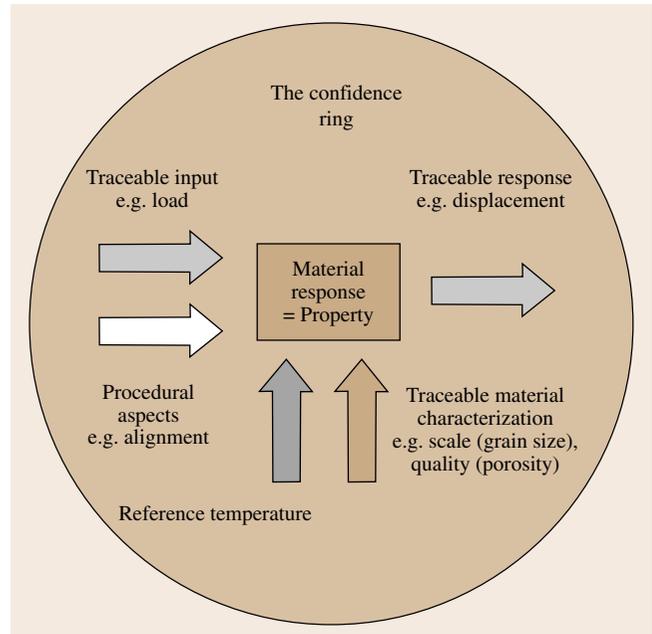


Fig. 1.4 Confidence ring for material property combined measurement and testing – note that separate traceability requirements apply to applied stimulus (load), response (displacement), and material characterization (grain size, porosity)

The confidence ring illustrates that, in measurement and testing, it is generally essential to establish reliable traceability for the applied stimulus and the resulting measured effect as well as for the measurements of any other quantities that may influence the final result. The final result may also be affected by the measurement procedure, by temperature, and by the state of the sample. It is important to understand that variation in measured results will often reflect material inhomogeneity as well as uncertainties associated with the test method or operator variability. All uncertainties should be taken into account in an *uncertainty budget*.

1.1.3 Conformity Assessment and Accreditation

In today's global market and world trade there is an increased need for *conformity assessment* to ensure that products and equipment meet specifications. The basis for conformity assessment are measurements together with methods of calibration, testing, inspection, and certification. The goal of conformity assessment

Table 1.1 Standards of conformity assessment tools

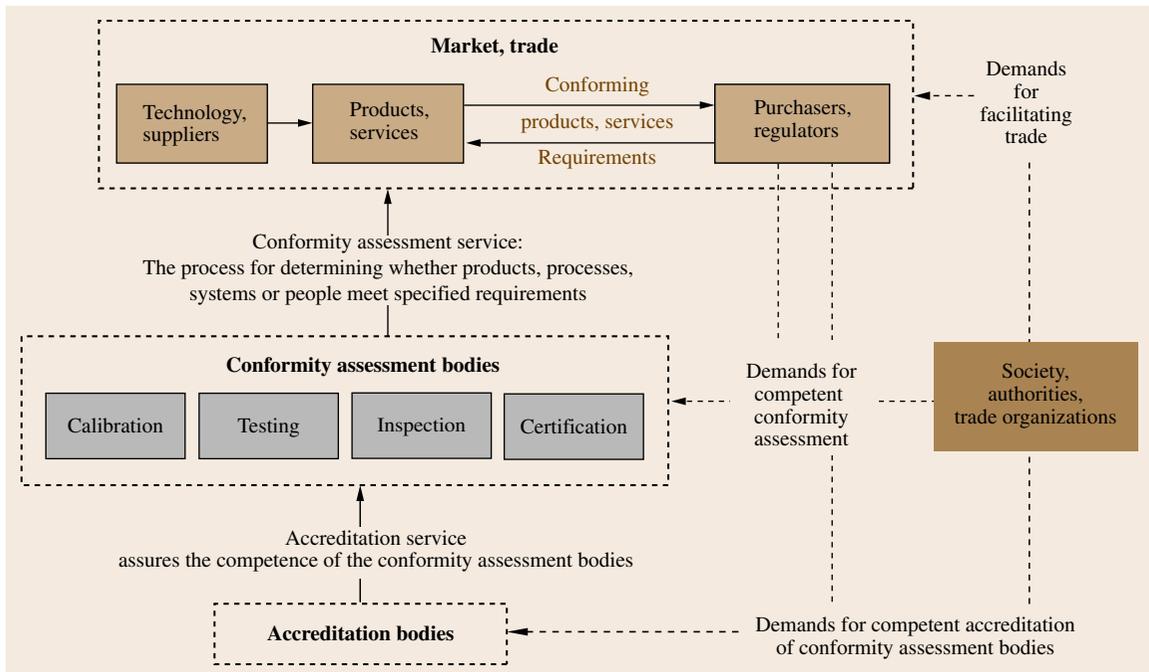
Tools for conformity assessment	First party Supplier, user	Second party Customers, trade associations, regulators	Third party Bodies independent from 1st and 2nd parties	ISO standards
Supplier's declaration	×			ISO/IEC 17050
Calibration, testing	×	×	×	ISO/IEC 17025
Inspection	×	×	×	ISO/IEC 17020
Certification			×	ISO 17021 ISO Guide 65

is to provide the user, purchaser or regulator with the necessary confidence that a product, service, process, system or person meets relevant requirements. The international standards relevant for conformity assessment services are provided by the ISO Committee on Conformity Assessment (CASCO). The conformity assessment tools are listed in Table 1.1, where their use by first parties (suppliers), second parties (customers, regulators, trade organizations), and third parties (bodies independent from both suppliers and customers) is indicated.

Along with the growing use of these conformity assessment tools there is the request for assurance of the competence of the conformity assessment bodies

(CABs). An increasingly applied and recognized tool for this assurance is *accreditation of CABs*.

The world's principal international forum for the development of laboratory accreditation practices and procedures is the International Laboratory Accreditation Cooperation (ILAC, <http://www.ilac.org/>). It promotes laboratory accreditation as a trade facilitation tool together with the recognition of competent calibration and testing facilities around the globe. ILAC started as a conference in 1977 and became a formal cooperation in 1996. In 2000, 36 ILAC members signed the ILAC Mutual Recognition Arrangement (MRA), and by 2008 the number of members of the ILAC MRA had risen to 60. Through the evaluation

**Fig. 1.5** Interrelations between market, trade, conformity assessment, and accreditation

of the participating accreditation bodies, the international acceptance of test data and the elimination of technical barriers to trade are enhanced as recommended and in support of the World Trade Organization

(WTO) Technical Barriers to Trade agreement. An overview of the interrelations between market, trade, conformity assessment, and accreditation is shown in Fig. 1.5.

1.2 Overview of Metrology

Having considered the methodologies of measurement and testing, a short general overview of metrology is given, based on *Metrology – in short* [1.5], a brochure published by EURAMET to establish a common metrological frame of reference.

1.2.1 The Meter Convention

In the middle of the 19th century the need for a worldwide decimal metric system became very apparent, particularly during the first universal industrial exhibitions. In 1875, a diplomatic conference on the meter took place in Paris, at which 17 governments signed the diplomatic treaty *the Meter Convention*. The signatories

decided to create and finance a permanent scientific institute: the *Bureau International des Poids et Mesures* (BIPM). The Meter Convention, slightly modified in 1921, remains the basis of all international agreement on units of measurement. Figure 1.6 provides a brief overview of the Meter Convention Organization (details are described in Chap. 2).

1.2.2 Categories of Metrology

Metrology covers three main areas of activities [1.5].

1. The definition of internationally accepted units of measurement

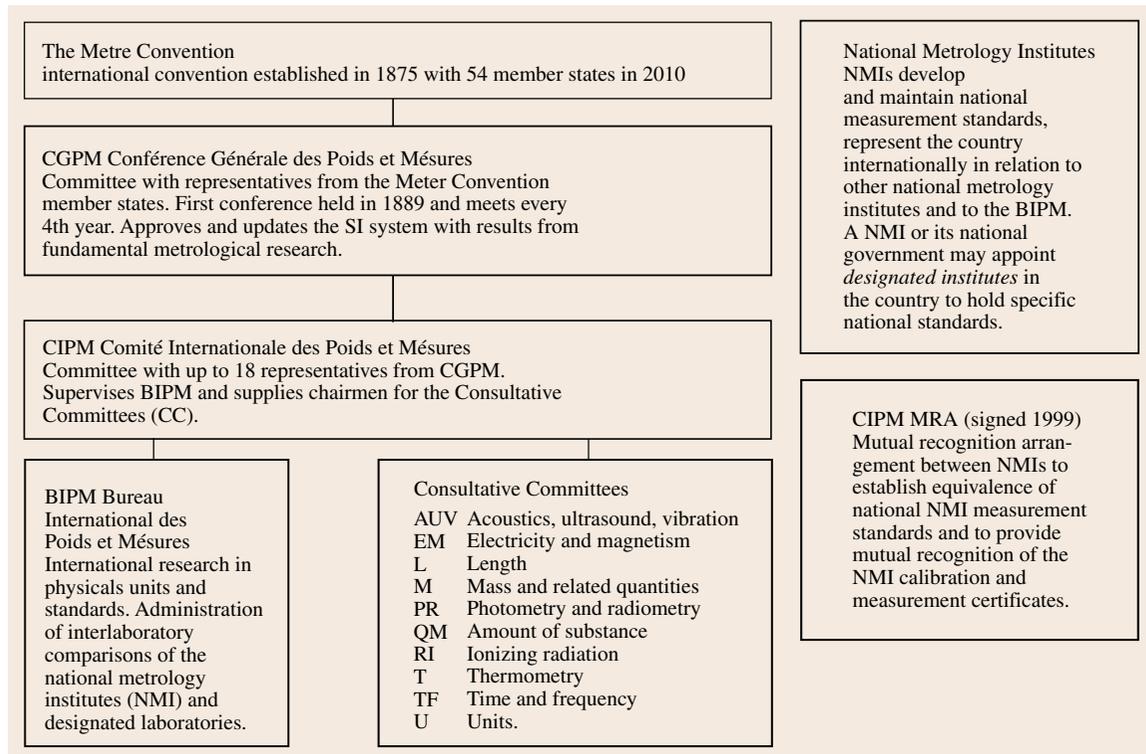


Fig. 1.6 The organizations and their relationships associated with the Meter Convention

2. The realization of units of measurement by scientific methods
3. The establishment of traceability chains by determining and documenting the value and accuracy of a measurement and disseminating that knowledge

Metrology is separated into three categories with different levels of complexity and accuracy (for details, see Chaps. 2 and 3).

Scientific Metrology

Scientific metrology deals with the organization and development of measurement standards and their maintenance. Fundamental metrology has no international definition, but it generally signifies the highest level of accuracy within a given field. Fundamental metrology may therefore be described as the top-level branch of scientific metrology.

Scientific metrology is categorized by BIPM into nine technical subject fields with different branches. The metrological calibration and measurement capabilities (CMCs) of the national metrology institutes (NIMs) and the designated institutes (DIs) are compiled together with *key comparisons* in the BIPM key comparison database (KCDB, <http://kcdb.bipm.org/>). All CMCs have undergone a process of peer evaluation by NMI experts under the supervision of the regional metrology organizations (RMOs). Table 1.2 shows the scientific metrology fields and their branches together with the number of registered calibration and measurement capabilities (CMCs) of the NMIs in 2010.

Industrial Metrology

Industrial metrology has to ensure the adequate functioning of measurement instruments used in industrial production and in testing processes. Systematic measurement with known degrees of uncertainty is one of the foundations of industrial quality control. Generally speaking, in most modern industries the costs bound up in taking measurements constitute 10–15% of production costs.

However, good measurements can significantly increase the value, effectiveness, and quality of a product. Thus, metrological activities, including calibration, testing, and measurements, are valuable inputs to ensure the quality of most industrial processes and quality of life related activities and processes. This includes the need to demonstrate traceability to international standards, which is becoming just as important as the measurement itself. Recognition of metrological competence at each level of the traceability chain can be established through mutual recognition agreements or arrangements, as well as through accreditation and peer review.

Legal Metrology

Legal metrology originated from the need to ensure fair trade, specifically in the area of weights and measures. The main objective of legal metrology is to assure citizens of correct measurement results when used in official and commercial transactions. Legally controlled instruments should guarantee correct measurement results throughout the whole period of use under working conditions, within given permissible errors.

Table 1.2 Metrology areas and their branches, together with the numbers of metrological calibration and measurement capabilities (CMCs) of the national metrology institutes and designated institutes in the BIPM KCDB as of 2010

Metrology area	Branch	CMCs
Acoustics, ultrasound, vibrations	Sound in air; sound in water; vibration	955
Electricity and magnetism	DC voltage, current, and resistance; impedance up to the megahertz range; AC voltage, current, and power; high voltage and current; other DC and low-frequency measurements; electric and magnetic fields; radiofrequency measurements	6586
Length	Laser; dimensional metrology	1164
Mass and related quantities	Mass; density; pressure; force; torque, viscosity, hardness and gravity; fluid flow	2609
Photometry and radiometry	Photometry; properties of detectors and sources; spectral properties; color; fiber optics	1044
Amount of substance	List of 16 amount-of-substance categories	4558
Ionizing radiation	Dosimetry; radioactivity; neutron measurements	3983
Thermometry	Temperature; humidity; thermophysical quantities	1393
Time and frequency	Time scale difference; frequency; time interval	586

For example, in Europe, the marketing and usage of the following measuring instruments are regulated by the European Union (EU) measuring instruments directive (MID 2004/22/EC)

1. Water meters
2. Gas meters
3. Electrical energy meters and measurement transformers
4. Heat meters
5. Measuring systems for liquids other than water
6. Weighing instruments
7. Taximeters
8. Material measures
9. Dimensional measuring systems
10. Exhaust gas analyzers

Member states of the European Union have the option to decide which of the instrument types they wish to regulate.

The International Organization of Legal Metrology (OIML) is an intergovernmental treaty organization established in 1955 on the basis of a convention, which was modified in 1968. In the year 2010, OIML was composed of 57 member countries and an additional 58 (corresponding) member countries that joined the OIML (<http://www.oiml.org/>) as observers. The purpose of OIML is to promote global harmonization of legal metrology procedures. The OIML has developed a worldwide technical structure that provides its members with metrological guidelines for the elaboration of national and regional requirements concerning the

manufacture and use of measuring instruments for legal metrology applications.

1.2.3 Metrological Units

The idea behind the metric system – a system of units based on the meter and the kilogram – arose during the French Revolution when two platinum artefact reference standards for the meter and the kilogram were constructed and deposited in the French National Archives in Paris in 1799 – later to be known as the Meter of the Archives and the Kilogram of the Archives. The French Academy of Science was commissioned by the National Assembly to design a new system of units for use throughout the world, and in 1946 the MKSA system (meter, kilogram, second, ampere) was accepted by the Meter Convention countries. The MKSA was extended in 1954 to include the kelvin and candela. The system then assumed the name the International System of Units (*Le Système International d'Unités*, SI). The SI system was established in 1960 by the 11th General Conference on Weights and Measures (CGPM): *The International System of Units (SI) is the coherent system of units adopted and recommended by the CGPM.*

At the 14th CGPM in 1971 the SI was again extended by the addition of the mole as base unit for amount of substance. The SI system is now comprised of seven base units, which together with derived units make up a coherent system of units [1.5], as shown in Table 1.3.

Table 1.3 The SI base units

Quantity	Base unit	Symbol	Definition
Length	Meter	m	The meter is the length of the path traveled by light in a vacuum during a time interval of $1/299\,792\,458$ of a second
Mass	Kilogram	kg	The kilogram is equal to the mass of the international prototype of the kilogram
Time	Second	s	The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom
Electric current	Ampere	A	The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed one meter apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} newtons per meter of length
Temperature	Kelvin	K	The kelvin is the fraction $1/273.16$ of the thermodynamic temperature of the triple point of water
Amount of substance	Mole	mol	The mole is the amount of substance of a system that contains as many elementary entities as there are atoms in 0.012 kg of carbon-12. When the mole is used, the elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles, or specified groups of such particles
Luminous intensity	Candela	cd	The candela is the luminous intensity in a given direction of a source that emits monochromatic radiation of frequency 540×10^{12} Hz and has a radiant intensity in that direction of $1/683$ W per steradian

Table 1.4 Examples of SI derived units expressed in SI base units

Derived quantity	SI derived unit special name	Symbol	In SI units	In SI base units
Force	Newton	N		m kg s^{-2}
Pressure, stress	Pascal	Pa	N/m^2	$\text{m}^{-1} \text{kg s}^{-2}$
Energy, work, quantity of heat	Joule	J	N m	$\text{m}^2 \text{kg s}^{-2}$
Power	Watt	W	J/s	$\text{m}^2 \text{kg s}^{-3}$
Electric charge	Coulomb	C		s A
Electromotive force	Volt	V		$\text{m}^2 \text{kg s}^{-3} \text{A}^{-1}$
Electric capacitance	Farad	F	C/V	$\text{m}^{-2} \text{kg}^{-1} \text{s}^4 \text{A}^2$
Electric resistance	Ohm	Ω	V/A	$\text{m}^2 \text{kg s}^{-3} \text{A}^{-2}$
Electric conductance	Siemens	S	A/V	$\text{m}^{-2} \text{kg}^{-1} \text{s}^3 \text{A}^2$

SI derived units are derived from the SI base units in accordance with the physical connection between the quantities. Some derived units, with examples from mechanical engineering and electrical engineering, are compiled in Table 1.4.

1.2.4 Measurement Standards

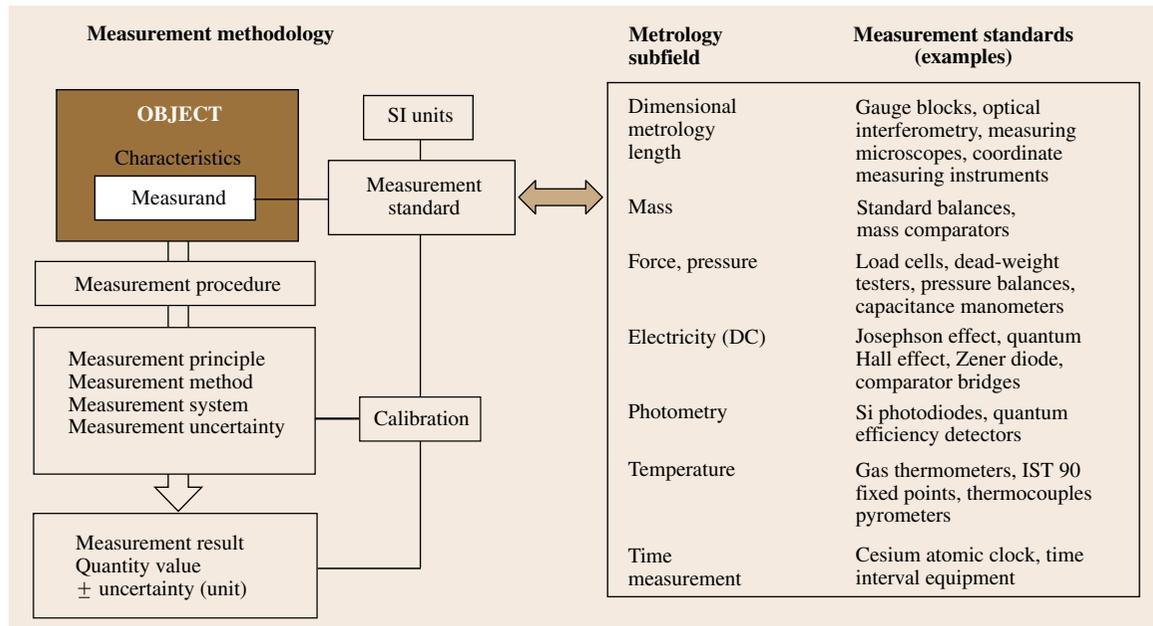
In the introductory explanation of the methodology of measurement, two essential aspects were pointed out.

1. Measurement begins with the definition of the measurand.

2. When the measurand is defined, it must be related to a measurement standard.

A measurement standard, or etalon, is the realization of the definition of a given quantity, with stated quantity value and associated measurement uncertainty, used as a reference. The realization may be provided by a material measure, measuring instrument, reference material or measuring system.

Typical measurement standards for subfields of metrology are shown in Fig. 1.7 in connection with the scheme of the measurement methodology (left-hand side of Fig. 1.1). Consider, for example, dimensional metrology. The meter is defined as the length of the path

**Fig. 1.7** Measurement standards as an integral part of the measurement methodology

traveled by light in vacuum during a time interval of $1/299\,792\,458$ of a second. The meter is realized at the primary level (SI units) in terms of the wavelength from an iodine-stabilized helium-neon laser. On sublevels, material measures such as gage blocks are used, and traceability is ensured by using optical interferometry to determine the length of the gage blocks with reference to the above-mentioned laser light wavelength.

1.3 Fundamentals of Materials Characterization

Materials characterization methods have a wide scope and impact for science, technology, economy, and society, as materials comprise all natural and synthetic substances and constitute the physical matter of engineered and manufactured products.

For materials there is a comprehensive spectrum of *materials measurands*. This is due to the broad variety of metallic, inorganic, organic, and composite materials, their different chemical and physical nature, and the manifold attributes which are related to materials with respect to composition, microstructure, scale, synthesis, physical and electrical properties, and applications. Some of these attributes can be expressed in a metrological sense as numbers, such as density; some are Boolean, such as the ability to be recycled or not; some, such as resistance to corrosion, may be expressed as a ranking (poor, adequate, good, for instance); and some can only be captured in text and images [1.14]. As background for materials characterization methods, which are treated in parts B, C, D of the handbook, namely

- Chemical and microstructural analysis
- Materials properties measurement
- Materials performance testing

the essential features of materials are outlined in the next sections [1.15].

1.3.1 Nature of Materials

Materials can be natural (biological) in origin or synthetically processed and manufactured. According to their chemical nature, they are broadly grouped traditionally into inorganic and organic materials.

The physical structure of materials can be crystalline or amorphous, as well as mixtures of both structures. Composites are combinations of materials assembled together to obtain properties superior to those of their single constituents. Composites (C) are

A national measurement standard is recognized by a national authority to serve in a state or economy as the basis for assigning quantity values to other measurement standards for the kind of quantity concerned. An international measurement standard is recognized by signatories to an international agreement and intended to serve worldwide, e.g., the international prototype of the kilogram.

classified according to the nature of their matrix: metal (MM), ceramic (CM) or polymer (PM) matrix composites, often designated as MMCs, CMCs, and PMCs, respectively. Figure 1.8 illustrates, with characteristic examples, the spectrum of materials between the categories natural, synthetic, inorganic, and organic.

From the view of *materials science*, the fundamental features of a solid material are as listed below.

- Material's atomic nature: the atomic elements of the Periodic Table which constitute the chemical composition of a material
- Material's atomic bonding: the type of cohesive electronic interactions between the atoms (or molecules) in a material, empirically categorized into the following basic classes.
 - Ionic bonds form between chemical elements with very different electron negativity (tendency

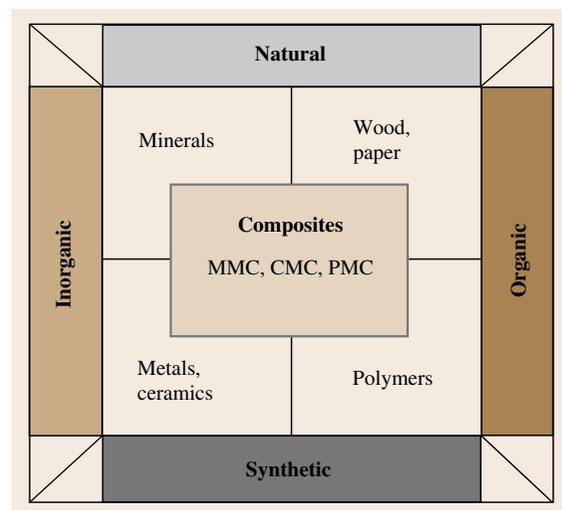


Fig. 1.8 Classification of materials

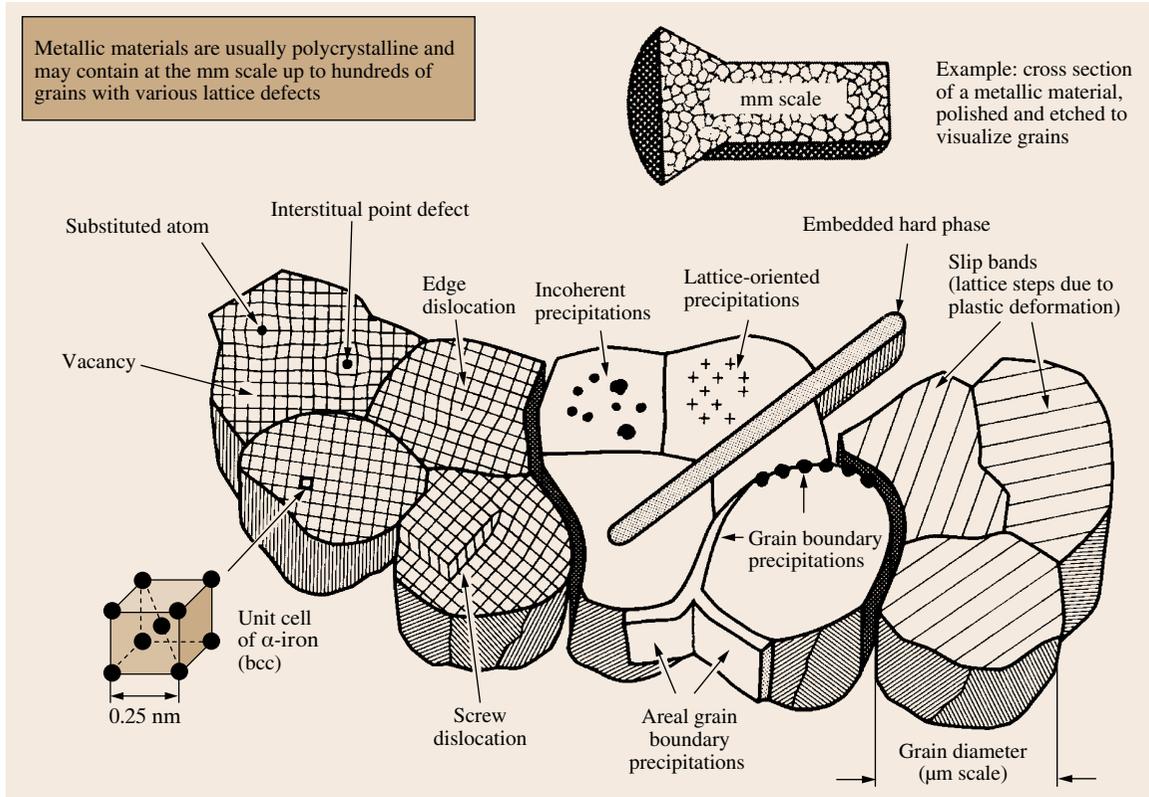


Fig. 1.9 Schematic overview on the microstructural features of metallic materials and alloys

to gain electrons), resulting in electron transfer and the formation of anions and cations. Bonding occurs through electrostatic forces between the ions.

- Covalent bonds form between elements that have similar electron negativities; the electrons are localized and shared equally between the atoms, leading to spatially directed angular bonds.
- Metallic bonds occur between elements with low electron negativities, so that the electrons are only loosely attracted to the ionic nuclei. A metal is thought of as a set of positively charged ions embedded in a sea of electrons.
- van der Waals bonds are due to the different internal electronic polarities between adjacent atoms or molecules, leading to weak (secondary) electrostatic dipole bonding forces.
- Material's spatial atomic structure: the amorphous or crystalline arrangement of atoms (or molecules) resulting from *long-range* or *short-range* bonding forces. In crystalline structures, it is characterized by unit cells which are the fundamental building blocks or modules, repeated many times in space within a crystal.
- Grains: crystallites made up of identical unit cells repeated in space, separated by grain boundaries.
- Phases: homogeneous aggregations of matter with respect to chemical composition and uniform crystal structure; grains composed of the same unit cells are the same phase.
- Lattice defects: deviations from ideal crystal structure.
 - Point defects or missing atoms: vacancies, interstitial or substituted atoms
 - Line defects or rows of missing atoms: dislocations
 - Area defects: grain boundaries, phase boundaries, and twins
 - Volume defects: cavities, precipitates.
- Microstructure: The microscopic collection of grains, phases, and lattice defects.

In addition to bulk materials characteristics, surface and interface phenomena also have to be considered.

In Fig. 1.9 an overview of the microstructural features of metallic materials is depicted schematically. Methods and techniques for the characterization of nanoscopic architecture and microstructure are presented in Chap. 5.

1.3.2 Types of Materials

It has been estimated that there are between 40 000 and 80 000 materials which are used or can be used in today's technology [1.14]. Figure 1.10 lists the main conventional *families* of materials together with examples of *classes*, *members*, and *attributes*. For the examples of attributes, necessary characterization methods are listed.

Metallic Materials and Alloys

In metals, the grains are the building blocks and are held together by the *electron gas*. The free valence electrons of the electron gas account for the high electrical and thermal conductivity, as well as for the optical gloss of metals. Metallic bonding, seen as the interaction between the total atomic nuclei and the electron gas, is not significantly influenced by displacement of atoms, which is the reason for the good ductility and formability of metals. Metals and metallic alloys are the most important group of the so-called *structural materials* whose special features for engineering applications are their mechanical properties, e.g., strength and toughness.

Semiconductors

Semiconductors have an intermediate position between metals and inorganic nonmetallic materials. Their most important representatives are the elements silicon and germanium, possessing covalent bonding and diamond structure; they are also similar in structure to III–V compounds such as gallium arsenide (GaAs). Being electric nonconductors at absolute zero temperature, semiconductors can be made conductive through thermal energy input or atomic doping, which leads to the creation of free electrons contributing to electrical conductivity. Semiconductors are important *functional materials* for electronic components and applications.

Inorganic Nonmetallic Materials or Ceramics

Atoms of these materials are held together by covalent and ionic bonding. As covalent and ionic bonding energies are much higher than those of metallic bonds, inorganic nonmetallic materials, such as ceramics, have high hardness and high melting temperatures. These materials are basically brittle and not ductile: In contrast to the metallic bond model, displacement of atomic dimensions theoretically breaks localized covalent bonds or transforms anion–cation attractions into anion–anion or cation–cation repulsions. Because of the lack of free valence electrons, inorganic nonmetallic materials are poor conductors of electricity and heat; this qualifies them as good insulators in engineering applications.

Organic Materials or Polymers and Blends

Organic materials, whose technologically most important representatives are the polymers, consist of macro-

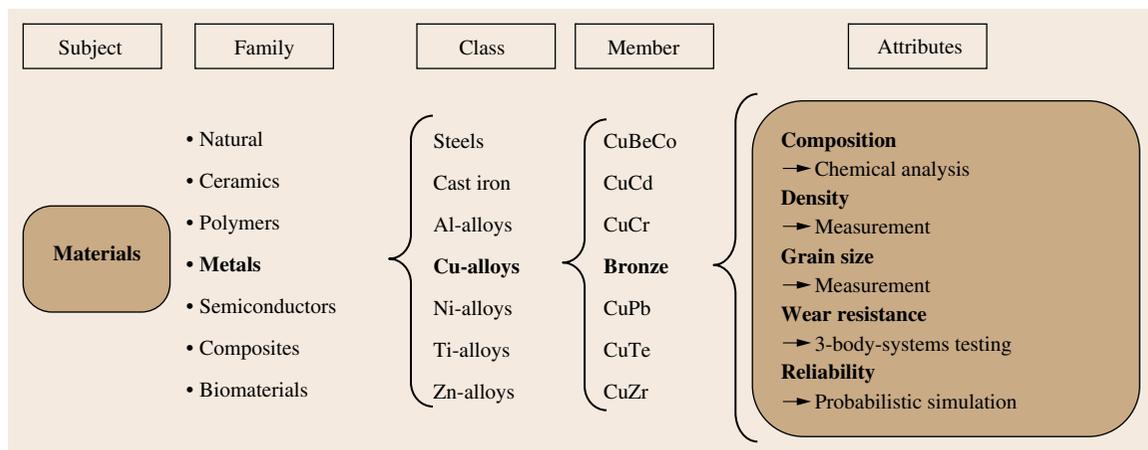


Fig. 1.10 Hierarchy of materials, and examples of attributes and necessary characterization methods

molecules containing carbon (C) covalently bonded with itself and with elements of low atomic number (e.g., H, N, O, S). Intimate mechanical mixtures of several polymers are called blends. In thermoplastic materials, the molecular chains have long linear structures and are held together by (weak) intermolecular (van der Waals) bonds, leading to low melting temperatures. In thermosetting materials, the chains are connected in a network structure and therefore do not melt. Amorphous polymer structures (e.g., polystyrene) are transparent, whereas crystalline polymers are translucent to opaque. The low density of polymers gives them a good strength-to-weight ratio and makes them competitive with metals in structural engineering applications.

Composites

Generally speaking, composites are hybrid creations made of two or more materials that maintain their identities when combined. The materials are chosen so that the properties of one constituent enhance the deficient properties of the other. Usually, a given property of a composite lies between the values for each constituent, but not always. Sometimes, the property of a composite is clearly superior to those of either of the constituents. The potential for such a synergy is one reason for the interest in composites for high-performance applications. However, because manufacturing of composites involves many steps and is labor intensive, composites may be too expensive to compete with metals and polymers, even if their properties are superior. In high-technology applications of advanced composites, it should also be borne in mind that they are usually difficult to recycle.

Natural Materials

Natural materials used in engineering applications are classified into natural materials of mineral origin, e.g., marble, granite, sandstone, mica, sapphire, ruby, or diamond, and those of organic origin, e.g., timber, India rubber, or natural fibres such as cotton and wool. The properties of natural materials of mineral origin, for example, high hardness and good chemical durability, are determined by strong covalent and ionic bonds between their atomic or molecular constituents and stable crystal structures. Natural materials of organic origin often possess complex structures with directionally dependent properties. Advantageous aspects of natural materials are ease of recycling and sustainability.

Biomaterials

Biomaterials can be broadly defined as the class of materials suitable for biomedical applications. They may be synthetically derived from nonbiological or even inorganic materials, or they may originate in living tissues. Products that incorporate biomaterials are extremely varied and include artificial organs; biochemical sensors; disposable materials and commodities; drug-delivery systems; dental, plastic surgery, ear, and ophthalmological devices; orthopedic replacements; wound management aids; and packaging materials for biomedical and hygienic uses. When applying biomaterials, understanding of the interactions between synthetic substrates and biological tissues is of crucial importance to meet clinical requirements.

1.3.3 Scale of Materials

The geometric length scale of materials covers more than 12 orders of magnitude. The scale ranges from

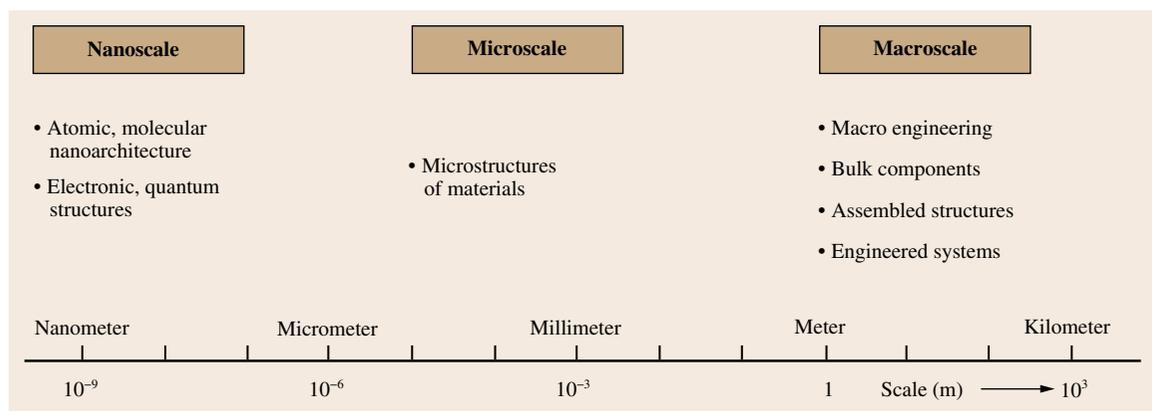


Fig. 1.11 Scale of material dimensions to be recognized in materials metrology and testing

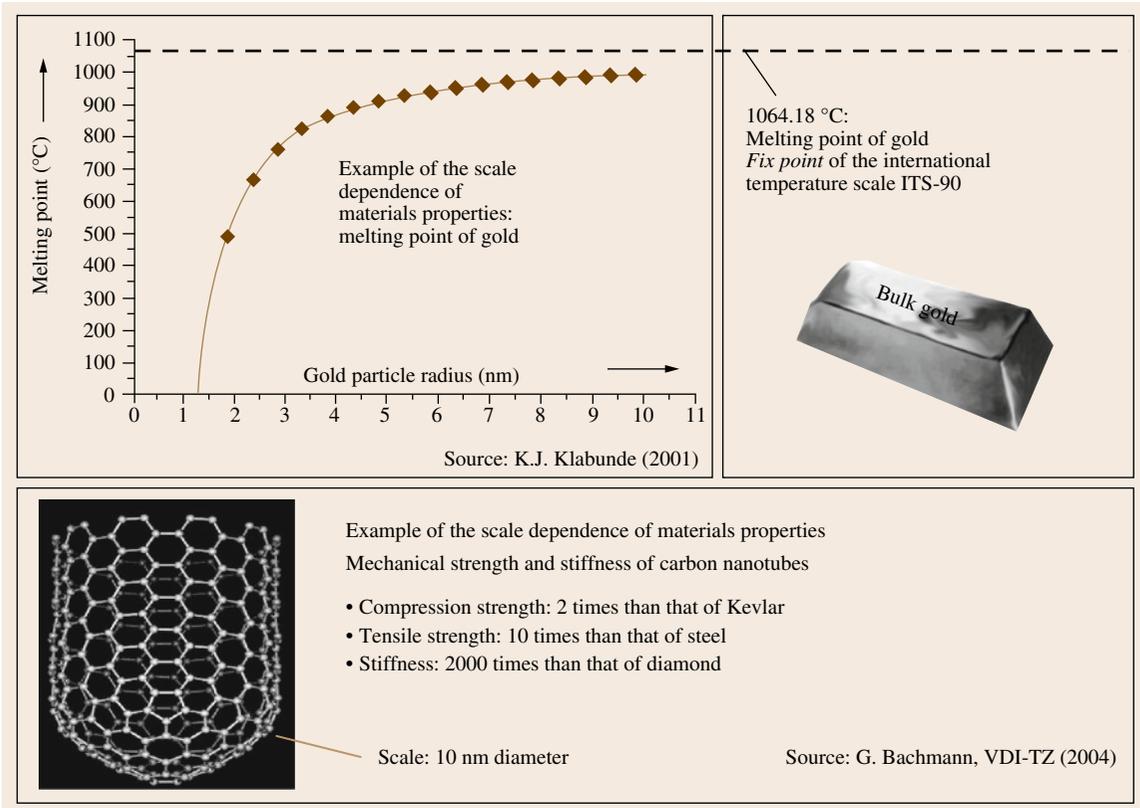


Fig. 1.12 Examples of the influence of scale effects on thermal and mechanical materials properties

the nanoscopic materials architecture to kilometer-long structures of bridges for public transport, pipelines, and oil drilling platforms for supplying energy to society. Figure 1.11 illustrates the dimensional scales relevant for today's materials science and technology.

Material specimens of different geometric dimensions have different bulk-to-surface ratios and may also have different bulk and surface microstructures. This can significantly influence the properties of materials, as exemplified in Fig. 1.12 for thermal and mechanical properties. Thus, scale effects have to be meticulously considered in materials metrology and testing.

1.3.4 Properties of Materials

Materials and their characteristics result from the processing of matter. Their properties are the response to extrinsic loading in their application. For every application, materials have to be engineered by processing, manufacturing, machining, forming or nanotechnology assembly to create structural, func-

tional or smart materials for the various engineering tasks (Fig. 1.13).

The properties of materials, which are of fundamental importance for their engineering applications, can be categorized into three basic groups.

1. Structural materials have specific mechanical or thermal properties for mechanical or thermal tasks in engineering structures.
2. Functional materials have specific electromagnetic or optical properties for electrical, magnetic or optical tasks in engineering functions.
3. Smart materials are engineered materials with intrinsic or embedded sensor and actuator functions, which are able to accommodate materials in response to external loading, with the aim of optimizing material behavior according to given requirements for materials performance.

Numerical values for the various materials properties can vary over several orders of magnitude for the different material types. An overview of the broad

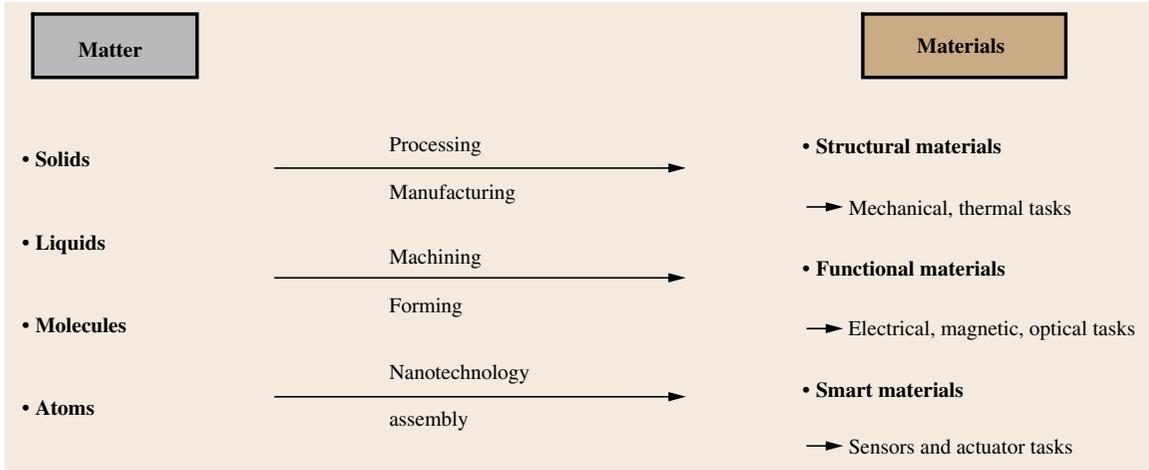


Fig. 1.13 Materials and their characteristics result from the processing of matter

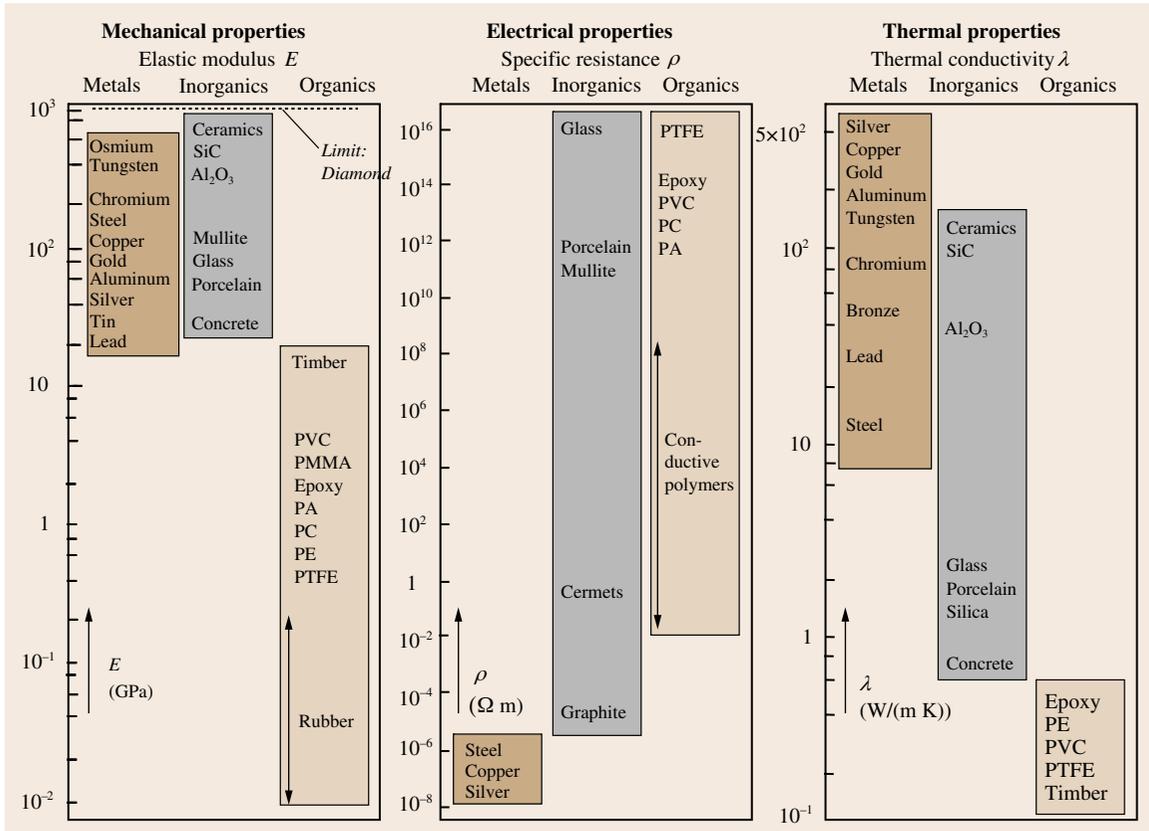


Fig. 1.14 Overview of mechanical, electrical, and thermal materials properties for the basic types of materials (metal, inorganic, or organic)

numerical spectra of some mechanical, electrical, and thermal properties of metals, inorganics, and organics is shown in Fig. 1.14 [1.16].

It must be emphasized that the numerical ranking of materials in Fig. 1.14 is based on *rough, average values* only. Precise data of materials properties require the specification of various influencing factors described above and symbolically expressed as

$$\begin{aligned} &\text{Materials properties data} \\ &= f(\text{composition–microstructure–scale,} \\ &\quad \text{external loading, } \dots) . \end{aligned}$$

1.3.5 Performance of Materials

For the application of materials as constituents of engineered products, performance characteristics such as quality, reliability, and safety are of special importance. This adds performance control and material failure analysis to the tasks of application-oriented materials measurement, testing, and assessment. Because all materials interact with their environment, materials–environment interactions and detrimental influences on

the integrity of materials must also be considered. An overview of the manifold aspects to be recognized in the characterization of materials performance is provided in Fig. 1.15.

The so-called materials cycle depicted schematically in Fig. 1.15 applies to all manmade technical products in all branches of technology and economy. The materials cycle illustrates that materials (accompanied by the necessary flow of energy and information) move in *cycles* through the technoeconomic system: from raw materials to engineering materials and technical products, and finally, after the termination of their task and performance, to deposition or recycling.

The operating conditions and influencing factors for the performance of a material in a given application stem from its structural tasks and functional loads, as shown in the right part of Fig. 1.15. In each application, materials have to fulfil technical functions as constituents of engineered products or parts of technical systems. They have to bear mechanical stresses and are in contact with other solid bodies, aggressive gases, liquids or biological species. In their functional tasks, materials always interact with their environment,

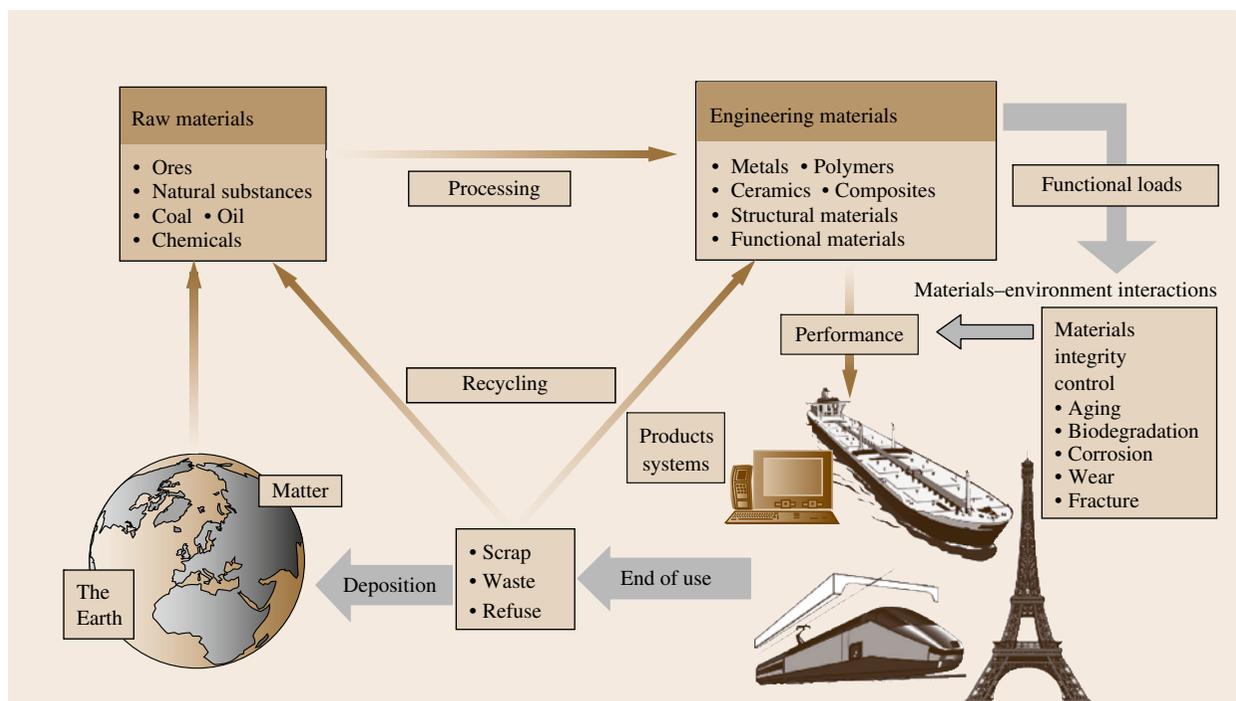


Fig. 1.15 The materials cycle of all products and technical systems

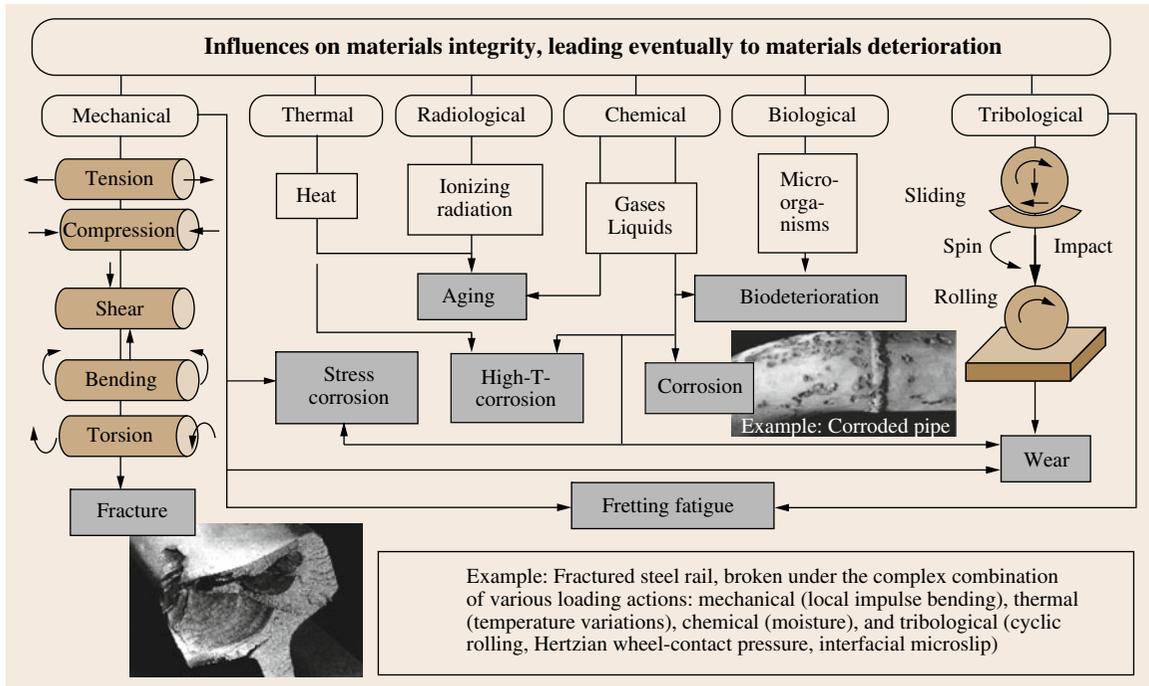


Fig. 1.16 Fundamentals of materials performance characterization: influencing phenomena

so these aspects also have to be recognized to characterize materials performance.

For the proper performance of engineered materials, materials deterioration processes and potential failures, such as materials aging, biodegradation, corrosion, wear, and fracture, must be controlled. Figure 1.16 shows an overview of influences on materials integrity and possible failure modes.

Figure 1.16 illustrates in a generalized, simplified manner that the influences on the integrity of materials, which are essential for their performance, can be categorized in mechanical, thermal, radiological, chemical, biological, and tribological terms. The basic materials deterioration mechanisms, as listed in Fig. 1.15, are aging, biodegradation, corrosion, wear, and fracture.

The deterioration and failure modes illustrated in Fig. 1.16 are of different relevance for the two elementary classes of materials, namely organic materials and inorganic materials (Fig. 1.8). Whereas aging and biodegradation are main deterioration mechanisms for organic materials such as polymers, the various types of corrosion are prevailing failure modes of metallic materials. Wear and fracture are relevant as materials deterioration and failure mechanisms for all types of materials.

1.3.6 Metrology of Materials

The topics of measurement and testing applied to materials (in short *metrology of materials*) concern the accurate and fit-for-purpose determination of the behavior of a material throughout its lifecycle.

Recognizing the need for a sound technical basis for drafting codes of practice and specifications for advanced materials, the governments of countries of the Economic Summit (G7) and the European Commission signed a Memorandum of Understanding in 1982 to establish the Versailles Project on Advanced Materials and Standards (VAMAS, <http://www.vamas.org/>). This project supports international trade by enabling scientific collaboration as a precursor to the drafting of standards. Following a suggestion of VAMAS, the Comité International des Poids et Mesures (CIPM, Fig. 1.6) established an ad hoc Working Group on the Metrology Applicable to the Measurement of Material Properties. The findings and conclusions of the Working Group on Materials Metrology were published in a special issue of *Metrologia* [1.17]. One important finding is the *confidence ring for traceability in materials metrology* (Fig. 1.4).

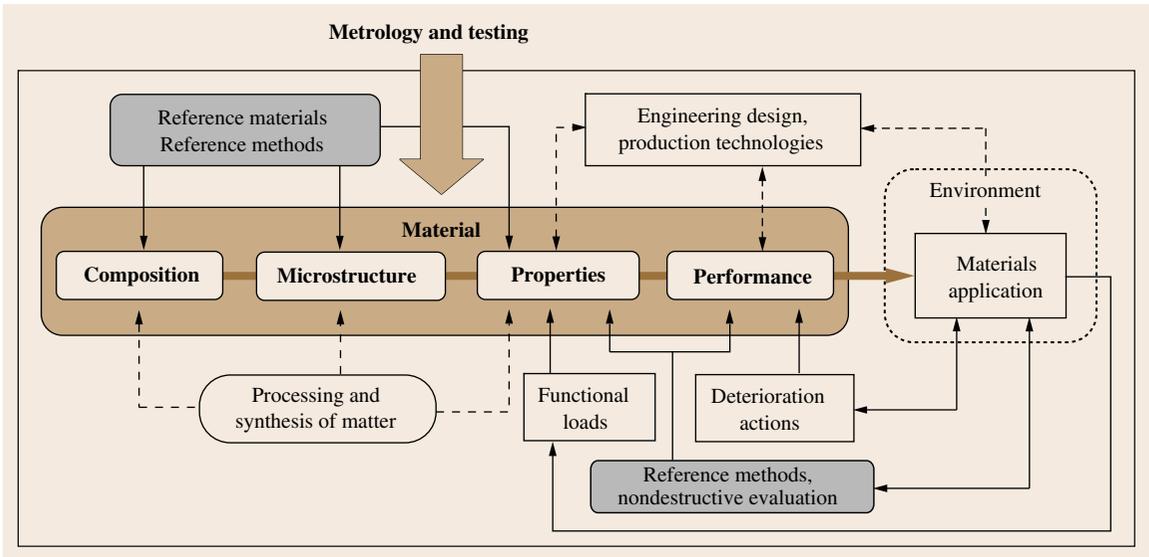


Fig. 1.17 Characteristics of materials to be recognized in metrology and testing

Materials in engineering design have to meet one or more structural, functional (e.g., electrical, optical, magnetic) or decorative purposes. This encompasses materials such as metals, ceramics, and polymers, resulting from the processing and synthesis of matter, based on chemistry, solid-state physics, and surface physics. Whenever a material is being created, developed or produced, the properties or phenomena that the material exhibits are of central concern. Experience shows that the properties and performance associated with a material are intimately related to its composition and structure at all scale levels, and influenced also by the engineering component design and production technologies. The final material, as a constituent of an engineered component, must perform a given task and must do so in an economical and societally acceptable manner. All these aspects are compiled in Fig. 1.17 [1.15].

The basic groups of materials characteristics essentially relevant for materials metrology and testing, as shown in the central part of Fig. 1.17, can be categorized as follows.

- *Intrinsic characteristics* are the material's composition and material's microstructure, described in Sect. 1.3.1. The intrinsic (inherent) materials characteristics result from the processing and syn-

thesis of matter. Metrology and testing to determine these characteristics have to be backed up by suitable reference materials and reference methods, if available.

- *Extrinsic characteristics* are the material's properties and material's performance, outlined in Sects. 1.3.4 and 1.3.5. They are *procedural characteristics* and describe the response of materials and engineered components to functional loads and environmental deterioration of the material's integrity. Metrology and testing to determine these characteristics have to be backed up by suitable reference methods and nondestructive evaluation (NDE).

It follows that, in engineering applications of materials, methods and techniques are needed to characterize intrinsic and extrinsic material attributes, and to consider also structure–property relations. The methods and techniques to characterize composition and microstructure are treated in part B of the handbook. The methods and techniques to characterize properties and performance are treated in parts C and D. The final part E of the handbook presents important modeling and simulation methods that underline measurement procedures that rely on mathematical models to interpret complex experiments or to estimate properties that cannot be measured directly.

References

- 1.1 BIPM: *International Vocabulary of Metrology*, 3rd edn. (BIPM, Paris 2008), available from <http://www.bipm.org/en/publications/guides/>
- 1.2 C. Ehrlich, S. Rasberry: Metrological timelines in traceability, *J. Res. Natl. Inst. Stand. Technol.* **103**, 93–106 (1998)
- 1.3 EUROLAB: Measurement uncertainty revisited: Alternative approaches to uncertainty evaluation, EUROLAB Tech. Rep. No. 1/2007 (EUROLAB, Paris 2007), <http://www.eurolab.org/>
- 1.4 ISO: *Guide to the Expression of Uncertainty in Measurement (GUM-1989)* (International Organization for Standardization, Geneva 1995)
- 1.5 P. Howarth, F. Redgrave: *Metrology – In Short*, 3rd edn. (Euramet, Braunschweig 2008)
- 1.6 ISO Guide 30: *Terms and Definitions of Reference Materials* (International Organization for Standardization, Geneva 1992)
- 1.7 T. Adam: *Guide for the Estimation of Measurement Uncertainty in Testing* (American Association for Laboratory Accreditation (A2LA), Frederick 2002)
- 1.8 EUROLAB: Guide to the evaluation of measurement uncertainty for quantitative tests results, EUROLAB Tech. Rep. No. 1/2006 (EUROLAB, Paris 2006), <http://www.eurolab.org/>
- 1.9 EURACHEM/CITAC Guide: *Quantifying Uncertainty in Analytical Measurement*, 2nd edn. (EURACHEM/CITAC, Lisbon 2000)
- 1.10 B. Magnusson, T. Naykki, H. Hovind, M. Krysell: *Handbook for Calculation of Measurement Uncertainty*, NORDTEST Rep. TR 537 (Nordtest, Espoo 2003)
- 1.11 R. Cook: *Assessment of Uncertainties of Measurement for Calibration and Testing Laboratories* (National Association of Testing Authorities Australia, Rhodes 2002)
- 1.12 J. Ma, I. Karaman: Expanding the repertoire of shape memory alloys, *Science* **327**, 1468–1469 (2010)
- 1.13 S. Bennett, G. Sims: Evolving needs for metrology in material property measurements – The conclusions of the CIPM Working Group on Materials Metrology, *Metrologia* **47**, 1–17 (2010)
- 1.14 M.F. Ashby, Y.J.M. Brechet, D. Cebon, L. Salvo: Selection strategies for materials and processes, *Mater. Des.* **25**, 51–67 (2004)
- 1.15 H. Czichos: Metrology and testing in materials science and engineering, *Measure* **4**, 46–77 (2009)
- 1.16 H. Czichos, B. Skrotzki, F.-G. Simon: Materials. In: *HÜTTE – Das Ingenieurwissen*, 33rd edn., ed. by H. Czichos, M. Hennecke (Springer, Berlin, Heidelberg 2008)
- 1.17 S. Bennett, J. Valdés (Eds.): Materials metrology, Foreword, *Metrologia* **47**(2) (2010)

2. Metrology Principles and Organization

This chapter describes the basic elements of metrology, the system that allows measurements made in different laboratories to be confidently compared. As the aim of this chapter is to give an overview of the whole field, the development of metrology from its roots to the birth of the Metre Convention and metrology in the 21st century is given.

2.1	The Roots and Evolution of Metrology	23	2.5	Regional Metrology Organizations	29
2.2	BIPM: The Birth of the Metre Convention	25	2.6	Metrological Traceability	29
2.3	BIPM: The First 75 Years	26	2.7	Mutual Recognition of NMI Standards: The CIPM MRA	30
2.4	Quantum Standards: A Metrological Revolution	28	2.7.1	The Essential Points of the MRA.....	30
			2.7.2	The Key Comparison Database (KCDB).....	31
			2.7.3	Take Up of the CIPM MRA	31
			2.8	Metrology in the 21st Century	32
			2.8.1	Industrial Challenges	32
			2.8.2	Chemistry, Pharmacy, and Medicine	33
			2.8.3	Environment, Public Services, and Infrastructures.....	34
			2.9	The SI System and New Science	34
			References	37	

2.1 The Roots and Evolution of Metrology

From the earliest times it has been important to compare things through measurements. This had much to do with fair exchange, barter, or trade between communities, and simple weights such as the stone or measures such as the cubit were common. At this level, parts of the body such as hands and arms were adequate for most needs. Initially, wooden length bars were easy to compare and weights could be weighed against each other. Various forms of balance were commonplace in early history and in religion. Egyptian tomb paintings show the Egyptian god Anubis weighing the soul of the dead against an ostrich feather – the sign of purity (Fig. 2.1). Noah's Ark was, so the Book of Genesis reports, 300 cubits long by 50 cubits wide and 30 cubits high. No one really knows why it was important to record such details, but the Bible, as just one example, is littered with metrological references, and the symbolism of metrology was part of early culture and art.

A steady progression from basic artifacts to naturally occurring reference standards has been part of the entire history of metrology. Metrologists are familiar with the use of the carob seed in early Mediterranean

civilizations as a natural reference for length and for weight and hence volume. The Greeks were early traders who paid attention to metrology, and they were known to keep copies of the weights and measures of the countries with which they traded.

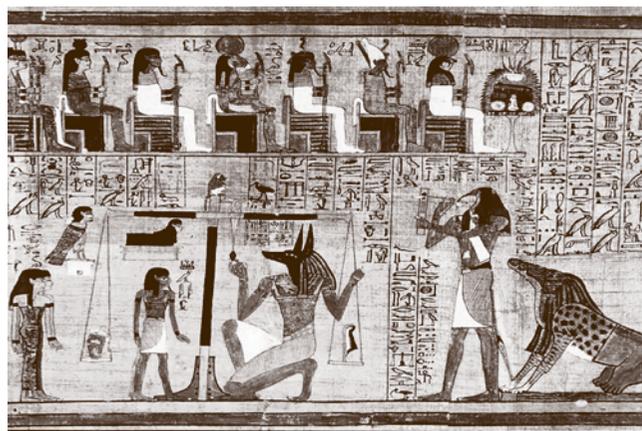


Fig. 2.1 Anubis

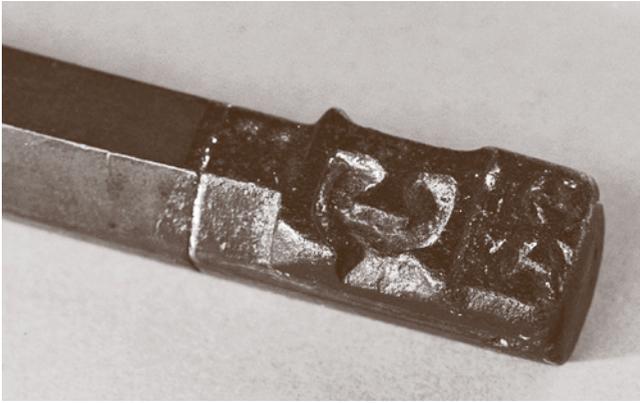


Fig. 2.2 Winchester yard



Fig. 2.3 Imperial length standards, Trafalgar Square, London

The Magna Carta of England set out a framework for a citizen's rights and established *one measure throughout the land*. Kings and queens took interest in national weights and measures; Fig. 2.2 shows the Winchester yard, the bronze rod that was the British standard from 1497 to the end of the 16th century. The queen's mark we see here is that of Elizabeth the First (1558–1603). In those days, the acre was widely used as a measurement, derived from the area that a team of oxen could plow in a day. Plowing an acre meant that you had to walk 66 furlongs, a linear dimension measured in rods.

The problem with many length measurements was that the standard was made of the commonly available metals brass or bronze, which have a fairly large coefficient of expansion. Iron or steel measures were not developed for a couple of centuries. Brass and bronze therefore dominated the length reference business until the early 19th century, by which time metallurgy had developed enough – though it was still something of a black art – for new, lower-expansion metals to be used and reference temperatures quoted. The UK imperial

references were introduced in 1854 and reproduced in Trafalgar Square in the center of London (Fig. 2.3), and the British Empire's domination of international commerce saw the British measurement system adopted in many of its colonies.

In the mid 18th century, Britain and France compared their national measurement standards and realized that they differed by a few percent for the same unit. Although the British system was reasonably consistent throughout the country, the French found that differences of up to 50% were common in length measurement within France. The ensuing technical debate was the start of what we now accept as the metric system. France led the way, and even in the middle of the French Revolution the Academy of Science was asked to *deduce an invariable standard for all the measures and all the weights*. The important word was *invariable*. There were two options available for the standard of length: the second's pendulum and the length of the Earth's meridian. The obvious weakness of the pendulum approach was that its period depended on the local acceleration due to gravity. The academy therefore chose to measure a portion of the Earth's circumference and relate it to the official French meter. Delambre and Méchain, who were entrusted with a survey of the meridian between Dunkirk and Barcelona, created the famous *Mètre des Archives*, a platinum end standard.

The international nature of measurement was driven, just like that of the Greek traders of nearly 2000 years before, by the need for interoperability in trade and advances in engineering measurement. The displays of brilliant engineering at the Great Exhibitions in London and Paris in the 19th century largely rested on the ability to measure well. The British Victorian engineer Joseph Whitworth coined the famous phrase *you can only make as well as you can measure* and pioneered accurate screw threads. He immediately saw the potential of end standard gages rather than line standards where the reference length was defined by a scratch on the surface of a bar. The difficulty, he realized, with line standards was that optical microscopes were not then good enough to compare measurements of line standards well enough for the best precision engineering. Whitworth was determined to make the world's best measuring machine. He constructed the quite remarkable *millionth machine* based on the principle that touch was better than sight for precision measurement. It appears that the machine had a *feel* of about 1/10 of a thousandth of an inch. Another interesting metrological trend at the 1851 Great Exhibition was military: the word *calibrate* comes from the need to control the caliber of guns.

At the turn of the 19th century, many of the industrialized countries had set up national metrology institutes, generally based on the model established in Germany with the Physikalisch Technische Reichsanstalt, which was founded in 1887. The economic benefits of such a national institute were immediately recognized, and as a result, scientific and industrial organizations in a number of industrialized countries began pressing their governments to make similar investments. In the UK, the British Association for the Advancement of Science reported that, without a national laboratory to act as a focus for metrology, the country's industrial competitiveness would be weakened. The cause was taken up more widely, and the UK set up the National Physical Laboratory (NPL) in 1900. The USA created the National Bureau of Standards in 1901 as a result of similar industrial pressure. The major national metrology institutes (NMIs), however, had a dual role. In general, they were the main focus for national research programs on applied physics and engineering. Their scientific role in the development of units – what became the International System of Units, the SI – began to challenge the role of the universities in the measurement of fundamental constants. This was especially true after the development of quantum physics in the 1920s and 1930s. Most early NMIs therefore began with two major elements to their mission

- a requirement to satisfy industrial needs for accurate measurements, through standardization and verification of instruments; and
- determination of physical constants so as to improve and develop the SI system.

The new industries which emerged after the First World War made huge demands on metrology and, together with mass production and the beginnings of multinational production sites, raised new challenges which brought NMIs into direct contact with companies, so causing a close link to develop between them. At that time, and even up to the mid 1960s, nearly all calibrations and measurements that were necessary for industrial use were made in the NMIs, and in the gage rooms of the major companies, as were most measurements in engineering metrology. The industries of the 1920s, however, developed a need for electrical and optical measurements, so NMIs expanded their coverage and their technical abilities.

The story since then is one of steady technical expansion until after the Second World War. In the 1950s, though, there was renewed interest in a broader applied focus for many NMIs so as to develop civilian applications for much of the declassified military technology. The squeeze on public budgets in the 1970s and 1980s saw a return to core metrology, and many other institutions, public and private, took on the responsibility for developing many of the technologies, which had been initially fostered at NMIs. The NMIs adjusted to their new roles. Many restructured and found new, often improved, ways of serving industrial needs. This recreation of the NMI role was also shared by most governments, which increasingly saw them as tools of industrial policy with a mission to stimulate industrial competitiveness and, at the end of the 20th century, to reduce technical barriers to world trade.

2.2 BIPM: The Birth of the Metre Convention

A brief overview of the Metre Convention has already been given in Sect. 1.2.1. In the following, the historical development will be described.

During the 1851 Great Exhibition and the 1860 meeting of the British Association for the Advancement of Science (BAAS), a number of scientists and engineers met to develop the case for a single system of units based on the metric system. This built on the early initiative of Gauss to use the 1799 meter and kilogram in the Archives de la République, Paris and the second, as defined in astronomy, to create a coherent set of units for the physical sciences. In 1874, the three-dimensional CGS system, based on the centimeter, gram, and second, was launched by the BAAS.

However, the size of the electrical units in the CGS system were not particularly convenient and, in the 1880s, the BAAS and the International Electrotechnical Commission (IEC) approved a set of practical electrical units based on the ohm, the ampere, and the volt. Parallel to this attention to the units, a number of governments set up what was then called the Committee for Weights and Money, which in turn led to the 1870 meeting of the Commission Internationale du Mètre. Twenty-six countries accepted the invitation of the French Government to attend; however, only 16 were able to come as the Franco-Prussian war intervened, so the full committee did not meet until 1872. The result was the Metre Convention and the creation of the Bureau International



Fig. 2.4 Bureau International des Poids et Mesures, Sèvres

des Poids et Mesures (BIPM, International Bureau of Weights and Measures), in the old Pavillon de Breteuil at Sèvres (Fig. 2.4), as a permanent scientific agency supported by the signatories to the convention. As this required the support of governments at the highest level, the Metre Convention was not finally signed until 20 May 1875.

The BIPM's role was to *establish new metric standards, conserve the international prototypes* (then the meter and the kilogram) and *to carry out the comparisons necessary to assure the uniformity of measures throughout the world*. As an intergovernmental, diplomatic treaty organization, the BIPM was placed under the authority of the General Conference on Weights and Measures (CGPM). A committee of 18 scientific experts, the International Committee for Weights and Measures (CIPM), now supervises the running of the BIPM. The aim of the CGPM and the CIPM was, and still is, to assure the *international unification and development of the metric system*. The CGPM now meets every 4 years to review progress, receive reports from the CIPM on the running of the BIPM, and establish the operating budget of the BIPM, whereas the CIPM meets annually to supervise the BIPM's work.

When it was set up, the staff of the BIPM consisted of a director, two assistants, and the necessary number of employees. In essence, then, a handful of

people began to prepare and disseminate copies of the international prototypes of the meter and the kilogram to member states. About 30 copies of the meter and 40 copies of the prototype kilogram were distributed to member states by ballot. Once this was done, some thought that the job of the BIPM would simply be that of periodically checking (in the jargon, verifying) the national copies of these standards. This was a short-lived vision, as the early investigations immediately showed the importance of reliably measuring a range of quantities that influenced the performance of the international prototypes and their copies. As a result, a number of studies and projects were launched which dealt with the measurement of temperature, density, pressure, and a number of related quantities. The BIPM immediately became a research body, although this was not recognized formally until 1921.

Returning to the development of the SI, one of the early decisions of the CIPM was to modify the CGS system to base measurements on the meter, kilogram, and second – the MKS system. In 1901, Giorgi showed that it was possible to combine the MKS system with the practical electrical units to form a coherent four-dimensional system by adding an electrical unit and rewriting some of the equations of electromagnetism in the so-called rationalized form. In 1946, the CIPM approved a system based on the meter, kilogram, second, and ampere – the MKSA system. Recognizing the ampere as a base unit of the metric system in 1948, and adding, in 1954, units for thermodynamic temperature (the kelvin and luminous intensity (the candela), the 11th CGPM in 1960 coined the name *Système internationale d'Unités*, the SI. At the 14th CGPM in 1971, the present-day SI system was completed by adding the mole as the base unit for the amount of substance, bringing the total number of base units to seven. Using these base units, a hierarchy of derived units and quantities of the SI have been developed for most, if not all, measurements needed in today's society. A substantial treatment of the SI is to be found in the 8th edition of the SI brochure published by the BIPM [2.1].

2.3 BIPM: The First 75 Years

After its intervention in the initial development of the SI, the BIPM continued to develop fundamental metrological techniques in mass and length measurement but soon had to react to the metrological implications of

major developments in atomic physics and interferometry. In the early 1920s, Albert Michelson came to work at the BIPM and built an eponymous interferometer to measure the meter in terms of light from

the cadmium red line – an instrument which, although modified, did sterling service until the 1980s. In temperature measurement, the old hydrogen thermometer scale was replaced with a thermodynamic-based scale and a number of fixed points. After great debate, electrical standards were added to the work of the BIPM in the 1920s with the first international comparisons of resistance and voltage. In 1929, an electrical laboratory was added, and photometry arrived in 1939.

In these early days, it was clear that the BIPM needed to find a way of consulting and collaborating with the experts in the world's NMIs. The solution adopted is one which still exists and flourishes today. The best way of working was in face-to-face meetings, so the concept of a consultative committee to the CIPM was born. Members of the committee were drawn from experts active in the world's NMIs and met to deal with matters concerning the definitions of units and the techniques of comparison and calibration. The consultative committees are usually chaired by a member of the CIPM. Much information was shared, although for obvious logistical reasons, the meetings were not too frequent. Over the years, the need for new consultative committees grew in reaction to the expansion of metrology, and now 10 consultative committees exist, with over 25 working groups.

The CIPM is rightly cautious about establishing a new committee, but proposals for new ones are considered from time to time, usually after an initial survey through a working group. In the last 10 years, joint committees have been created to tackle issues such as the

international approach to the estimation of measurement uncertainties or to the establishment of a common vocabulary for metrology. Most of these joint committees bring the BIPM together with international or intergovernmental bodies such as the International Organization for Standardization (ISO), the International Laboratory Accreditation Cooperation (ILAC), and the IEC. As the work of the Metre Convention moves into areas other than its traditional activities in physics and engineering, joint committees are an excellent way of bringing the BIPM together with other bodies that bring specialist expertise – an example being the joint committee for laboratory medicine, established recently with the International Federation of Clinical Chemists and the ILAC.

The introduction of ionizing radiation standards to the work of the BIPM came when Marie Curie deposited her first radium standard at the BIPM in 1913. As a result of pressure, largely from the USSR delegation to the CGPM, the CIPM took the decision to deal with metrology in ionizing radiation.

In the mid 1960s, and at the time of the expansion into ionizing radiation, programs on laser length measurement were also started. These contributed greatly to the redefinition of the meter in 1983. The BIPM also acted as the world reference center for laser wavelength or frequency comparisons in much the same way as it did for physical artifact-based standards. In the meantime, however, the meter bar had already been replaced, in 1960, by an interferometric-based definition using optical radiation from a krypton lamp.

Table 2.1 List of consultative committees with dates of formation

Names of consultative committees and the date of their formation

(Names of consultative committees have changed as they have gained new responsibilities; the current name is cited)

Consultative Committee for Electricity and Magnetism, CCEM (1997, but created in 1927)
Consultative Committee for Photometry and Radiometry, CCPR (1971 but created in 1933)
Consultative Committee for Thermometry, CCT (1937)
Consultative Committee for Length, CCL (1997 but created in 1952)
Consultative Committee for Time and Frequency, CCTF (1997, but created 1956)
Consultative Committee for Ionizing Radiation, CCRI (1997, but created in 1958)
Consultative Committee for Units, CCU (1964, but replacing a similar commission created in 1954)
Consultative Committee for Mass and Related Quantities, CCM (1980)
Consultative Committee for Amount of Substance and Metrology in Chemistry, CCQM (1993)
Consultative Committee for Acoustics, Ultrasound, and Vibration, CCAUV (1999)

2.4 Quantum Standards: A Metrological Revolution

Technology did not stand still, and in reaction to the developments and metrological applications of superconductivity, important projects on the Josephson, quantum Hall, and capacitance standards were launched in the late 1980s and 1990s. The BIPM played a key role in establishing worldwide confidence in the performance of these new devices through an intense program of comparisons which revealed many of the systematic sources of error and found solutions to them. The emergence of these and other quantum-based standards, however, was an important and highly significant development. In retrospect, these were one of the drivers for change in the way in which world metrology organizes itself and had implications nationally as well as internationally.

The major technical change was, in essence, a belief that standards based on quantum phenomena were the same the world over. Their introduction sometimes made it unnecessary for a single, or a small number of, reference standards to be held at the BIPM or in a few of the well-established and experienced NMIs. Quantum-based standards were, in reality, available to all and, with care, could be operated outside the NMIs at very high levels of accuracy.

There were two consequences. Firstly, that the newer NMIs that wanted to invest in quantum-based standards needed to work with experienced metrologists in existing NMIs in order to develop their skills as rapidly as possible. Many of the older NMIs, therefore, became adept at training and providing the necessary experience for newer metrologists. The second consequence was an increased pressure for comparisons of standards, as the ever-conservative metrology community sought to develop confidence in new NMIs as well as in the quantum-based standards. These included frequency-stabilized lasers, superconducting voltage and resistance standards, cryogenic radiometers (for measurements related to the candela), atomic clocks (for the second), and a range of secondary standards.

Apart from its responsibility to maintain the international prototype kilogram (Fig. 2.5), which remains the last artifact-based unit of the SI, the BIPM was therefore no longer always the sole repository of an international primary reference standard. However there were, and still are, a number of unique reference facilities at the BIPM for secondary standards and quantities of the SI. Staff numbers had also leveled off at about 70.

If it was going to maintain its original mission of a scientifically based organization with responsibility

for coordinating world metrology, the BIPM recognized that it needed to discharge particular aspects of its treaty obligation in a different way. It also saw the increased value of developing the links needed to establish collaboration at the international and intergovernmental level. In addition, the staff had the responsibility to provide the secretariat to the 10 consultative committees of the CIPM as well as an increasing number of working groups. The last 10 years of the 20th century, therefore, saw the start of a significant change in the BIPM's way of working. During this period, it was also faced with the need to develop a world metrology infrastructure in new areas such as the environment, chemistry, medicine, and food. The shift away from physics and engineering was possible, fortunately, as a result of the changing way in which the SI units could be realized, particularly through the quantum-based standards. Other pressures for an increase in the BIPM's coordination role resulted from the increasingly intensive program of comparisons brought about by the launch of the CIPM's mutual recognition arrangement in the 1990s.

The most recent consequence of these trends was that the CIPM decided that the photometry and radiometry section would close due to the need to operate within internationally agreed budgets, ending nearly

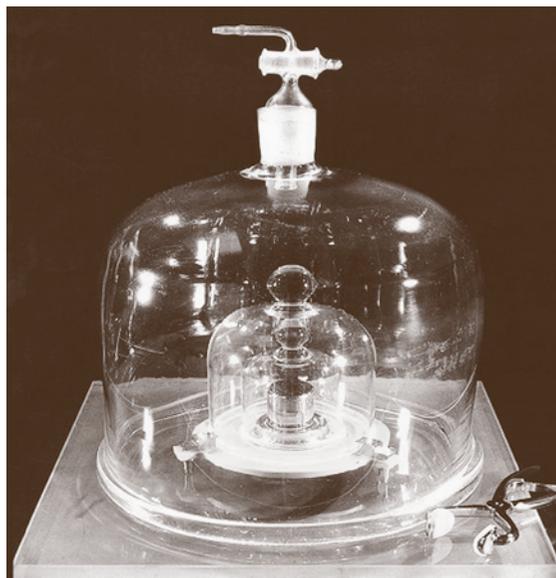


Fig. 2.5 International prototype kilogram (courtesy of BIPM)

70 years of scientific activity at the BIPM. Additional savings would also be made by restricting the work of the laser and length group to a less ambitious program.

A small, 130-year-old institution was therefore in the process of reinventing itself to take on and develop a changed but nevertheless unique niche role. This was still based on technical capabilities and laboratory work but was one which had to meet the changing, and expanding, requirements of its member states in a different

way. Much more also needed to be done as the benefits of precise, traceable measurement became seen as important in a number of the new disciplines for metrology. This change of emphasis was endorsed at the 2003 General Conference on Weights and Measures as a new 4 year work program (2005–2008) was agreed, as well as the first real-terms budget increase since the increase agreed in the mid 1960s which then financed the expansion into ionizing radiation.

2.5 Regional Metrology Organizations

The growth of the number of NMIs and the emergence of world economic groupings such as the Asia–Pacific Economic Cooperation and the European Union mean that regional metrological groupings have become a useful way of addressing specific regional needs and can act as a mutual help or support network. The first such group was probably in Europe, where an organization now named EURAMET emerged from a loose collaboration of NMIs based on the Western European Metrology Club. There are now five regional metrology organizations (RMOs): the Asian–Pacific Metrology Program (APMP) with perhaps the largest geographical coverage from India in the west to New Zealand in the east and extending into Asia, the Euro–Asian Cooperation in Metrology amongst the Central European Countries (COOMET), the European Association of National Metrology Institutes (EURAMET), the Southern African Development Community Cooperation in Measurement Traceability (SADC MET), and Sistema Interamericano de Metrología (SIM, Inter-American

Metrology System) which covers Southern, Central, and North America). The RMOs play a vital role in encouraging coherence within their region and between regions; without their help, the Metre Convention would be far more difficult to administer and its outreach to nonmembers – who may, however, be members of an RMO – would be more difficult.

Traditionally, NMIs have served their own national customers. It is only within the last 10 years that regional metrology organizations have started to become more than informal associations of national laboratories and have begun to develop strategies for mutual dependence and resource sharing, driven by concerns about budgets and the high cost of capital facilities. The sharing of resources is still, however, a relatively small proportion of all collaborations between NMIs, most of which are still at the research level. It is, of course, no coincidence that RMOs are based on economic or trading blocs and that these groupings are increasingly concerned with free trade within and between them.

2.6 Metrological Traceability

Traceability of measurement has been a core concern of the Metre Convention from its inception, as has been emphasized already in Sect. 2.1. Initially, a measurement is always made in relation to a more accurate reference, and these references are themselves calibrated or measured against an even more accurate reference standard. The chain follows the same pattern until one reaches the national standards. (For technical details see Chap. 3.)

The NMIs' job was to make sure that the national standards were traceable to the SI and were accurate enough to meet national needs. As NMIs themselves stopped doing all but the highest accuracy measurements and as accredited laboratories, usually

in the commercial sector, took on the more routine tasks, the concept of a national hierarchy of traceable measurements became commonplace, frequently called a national measurement system. In general, the technical capabilities of the intermediate laboratories are assured by their accreditation to the international documentary standards ISO/IEC 17025 by a national accreditation body, usually a member of the International Laboratory Accreditation Cooperation (ILAC) (Sect. 1.1.3). At the top of the traceability system, measurements were relatively few in number and had the lowest uncertainty. Progressing down the traceability chain introduced a greater level of uncertainty of measurement, and generally speaking, a larger number

of measurements are involved. Traceability itself also needed to be defined. The *international vocabulary of metrology (VIM)* defines traceability as [2.2]:

The property of a measurement result relating the result to a stated metrological reference through an unbroken chain of calibrations or comparisons each contributing to the stated uncertainty.

The important emphasis is on uncertainty of measurement (for a detailed treatment of measurement uncertainty, the reader is referred to the *Guide to the expression of uncertainty in measurement (GUM)* [2.3]) and the need for the continuous unbroken chain of measurement. Comparisons of standards or references are a common way of demonstrating confidence in the measurement processes and in the reference standards held either in the **NMIs** or in accredited laboratories. The national accreditation body usually takes care of these comparisons at working levels, sometimes called interlaboratory comparisons (**ILCs**) or proficiency testing (Sect. 3.6).

At the **NMI** level, the framework of the **BIPM** and the **CIPM**'s consultative committees (**CCs**) took care of the highest level comparisons. However, the increased relevance of traceable measurement to trade,

and the need for demonstrable equivalence of the national standards held at **NMIs**, and to which national measurements were traceable, took a major turn in the mid 1990s. This event was stimulated by the need, from the accreditation community as much as from regulators and trade bodies, to know just how well the **NMI** standards agreed with each other. Unlike much of the work of the consultative committees, this involved **NMIs** of all states of maturity working at all levels of accuracy. The task of comparing each and every standard was too great and too complex for the **CC** network, so a novel approach needed to be adopted. In addition, it became increasingly clear that the important concept was one of measurements traceable to the **SI** through the standards realized and maintained at **NMIs**, rather than to the **NMI**-based standards themselves. Not to develop and work with this concept would run the risk of creating technical barriers to trade (**TBTs**) if measurements in a certain country were legally required to be traceable to the **NMI** standards or if measurements made elsewhere were not recognized. The World Trade Organization was turning its attention towards the need for technical measurements to be accepted worldwide and was setting challenging targets for the reduction of **TBTs**. The metrology community needed to react.

2.7 Mutual Recognition of NMI Standards: The CIPM MRA

The result was the creation, by the **CIPM**, of a mutual recognition arrangement (**MRA**) for the recognition and acceptance of **NMI** calibration and test certificates. The **CIPM MRA** is one of the key events of the last few years, and one which may be as significant as the Metre Convention itself. The **CIPM MRA** has a direct impact on the reduction of technical barriers to trade and to the globalization of world business. The **CIPM MRA** was launched at a meeting of **NMIs** from member states of the Metre Convention held in Paris on 14 October 1999, at which the directors of the national metrology institutes of 38 member states of the convention and representatives of two international organizations became the first signatories.

2.7.1 The Essential Points of the MRA

The objectives of the **CIPM MRA** are

- to establish the degree of equivalence of national measurement standards maintained by **NMIs**,

- to provide for the mutual recognition of calibration and measurement certificates issued by **NMIs**, and
- thereby to provide governments and other parties with a secure technical foundation for wider agreements related to international trade, commerce, and regulatory affairs.

The procedure through which an **NMI**, or any other recognized signatory, joins the **MRA** is based on the need to demonstrate their technical competence, and to convince other signatories of their performance claims. In essence, these performance claims are the uncertainties associated with the routine calibration services which are offered to customers and which are traceable to the **SI**. Initial claims, called calibration and measurement capabilities (**CMCs**), are first made by the laboratory concerned. They are then first reviewed by technical experts from the local regional metrology organization and, subsequently, by other **RMOs**. The technical evidence for the **CMC** claims is generally based on the institute's performance in a number of comparisons

carried out and managed by the relevant CIPM consultative committees (CCs) or by the RMO. This apparently complex arrangement is needed because it would be technically, financially or organizationally impossible for each participant to compare its own SI standards with all others. The CIPM places particular importance on two types of comparisons

- international comparisons of measurements, known as CIPM key comparisons and organized by the CCs, which generally involve only those laboratories which perform at the highest level. The subject of a key comparison is chosen carefully by the CC to be representative of the ability of the laboratory to make a range of related measurements;
- key or supplementary international comparisons of measurements, usually organized by the RMOs and which include some of the laboratories which took part in the CIPM comparisons as well as other laboratories from the RMO. RMO key comparisons are in the same technical area as the CIPM comparison, whereas supplementary comparisons are usually carried out to meet a special regional need.

Using this arrangement, we can establish links between all participants to provide the technical basis for the comparability of the SI standards at each NMI. Reports of all the comparisons are published in the key comparison database maintained by the BIPM on its website (www.bipm.org).

These comparisons differ from those traditionally carried out by the CCs, which were largely for scientific reasons and which established the dependence of the SI realizations on the effects which contributed to the uncertainty of the realization. In CIPM and RMO key or supplementary comparisons, however, each participant carries out the measurements without knowing the results of others until the comparison has been completed. They provide, therefore, an independent assessment of performance. The CIPM, however, took the view that comparisons are made at a specific moment in time and so required participating NMIs to install a quality system which could help demonstrate confidence in the continued competence of participants in between comparisons. All participants have chosen to use the ISO/IEC 17025 standard and have the option of a third-party accreditation by an ILAC member or a self-declaration together with appropriate peer reviews.

The outcome of this process is that it gives NMIs the confidence to recognize the results of key and supplementary comparisons as stated in the database and

therefore to accept the calibration and measurement capabilities of other participating NMIs.

When drawing up its MRA, the CIPM was acutely aware that its very existence, and the mutual acceptance of test and calibration certificates between its members, might be seen as a technical barrier to trade in itself. The concept of associate members of the CGPM was therefore developed. An associate has, in general, the right to take part in the CIPM MRA but not to benefit from the full range of BIPM services and activities, which are restricted to convention members. Associate status is increasingly popular with developing countries, as it helps them gain recognition worldwide but does not commit them to the additional expense of convention membership, which may be less appropriate for them at their stage of development.

2.7.2 The Key Comparison Database (KCDB)

The key comparison database, referred to in the MRA and introduced in Sect. 1.2.2 (Table 1.2), is available on the BIPM website (www.bipm.org). The content of the database is evolving rapidly. Appendix A lists signatories, and appendix B contains details of the set of key comparisons together with the results from those that have been completed. The database will also contain a list of those old comparisons selected by the consultative committees that are to be used on a provisional basis. Appendix C contains the calibration and measurement capabilities of the NMIs that have already been declared and reviewed within their own regional metrology organization (RMO) as well as those other RMOs that support the MRA.

2.7.3 Take Up of the CIPM MRA

The KCDB data is, at the moment, largely of interest to metrologists. However, a number of NMIs are keen to see it taken up more widely, and there are several examples of references to the CIPM MRA in regulation. This campaign is at an early stage; an EU–USA trade agreement cites the CIPM MRA as providing an appropriate technical basis for acceptance of measurements and tests, and the USA's National Institute of Standards and Technology (NIST) has opened up a discussion with the Federal Aviation Authority (FAA) and other regulators as to the way in which they can use KCDB data to help the FAA accept the results of tests and certificates which have been issued outside the USA.

There is a range of additional benefits and consequences of the CIPM MRA. Firstly, anyone can use

the **KCDB** to look for themselves at the validated technical capability of any **NMI**. As a result, they can, with full confidence, choose to use its calibration services rather than those of their national laboratory and have the results of these services accepted worldwide. They can also use the **MRA** database to search for **NMIs** that can satisfy their needs if they are not available nationally. This easy access and the widespread and cheap availability of information may well drive a globalization of the calibration service market and will enable users to choose the supplier that best meets their needs. As the **MRA** is implemented, it will be a real test of market economics. Secondly, there is the issue of rapid turnarounds. Companies that have to send their standards away for calibration do not

have them available for in-house use. This can lead to costly duplication if continuity of an internal service is essential, or to a tendency to increase the calibration interval if calibrations are expensive. **NMIs** therefore have to concentrate increasingly on reducing turnaround times, or providing better customer information through calibration management systems. Some calibrations will always require reasonable periods of time away from the workplace because of the need for stability or because **NMIs** can only (through their own resource limitations) provide the service at certain times. This market sensitivity is now fast becoming built into service delivery and is, in some cases, more important to a customer than the actual price of a calibration.

2.8 Metrology in the 21st Century

In concluding this review of the work of the Metre Convention, it seems appropriate to take a glance at what the future may have in store for world metrology and, in particular, at the new industries and technologies which require new measurements.

2.8.1 Industrial Challenges

The success of the Metre Convention and, in particular, the recognized technical and economic benefits of the **CIPM MRA** – one estimate by the consulting company **KPMP** puts its potential impact on the reduction of **TBTs** at some €4 billion (see the **BIPM** website) – have attracted the interest and attention of new communities of users.

New Technologies

A challenge which tackles the needs of new industries and exploits new technologies is to enhance our ability to measure the large, the small, the fast, and the slow. Microelectronics, telecommunications, and the study and characterization of surfaces and thin films will benefit. Many of these trends are regularly analyzed by **NMIs** as they formulate their technical programs, and a summary can be found in the recent report to the 22nd **CGPM** by its secretary Dr. Robert Kaarls, entitled *Evolving Needs for Metrology in Trade, Industry, and Society and the Role of the BIPM*. The report can be found on the **BIPM** website (<http://www.bipm.org/>).

Materials Characterization

Of particular relevance to the general topic of this handbook, there has been an interest in a worldwide framework for traceable measurement of material characteristics. The potential need is for validated, accepted reference data to characterize materials with respect to their properties (see Part C) and their performance (see Part D). To a rather limited extent, some properties are already covered in a metrological sense (hardness, some thermophysical or optical properties, for example), but the vast bulk of materials characteristics, which are not intrinsic properties but system-dependent attributes (Sect. 1.3.6), remain outside the work of the convention. Therefore, relatively few **NMIs** have been active in materials metrology, but a group is meeting to decide whether to make proposals for a new sphere of Metre Convention activity. A report recommending action was presented to the **CIPM** in 2004 and will be followed up in a number of actions launched by the committee (Sect. 1.3.6).

Product Appearance

Another challenge is the focus by manufacturers on the design or appearance of a product, which differentiates it in the eyes of the consumer from those of their competitors. These rather subjective attributes of a product are starting to demand an objective basis for comparison. Appearance measurement of quantities such as gloss, or the need to measure the best conditions in which to display products under different lighting or pre-

sentation media (such as a TV tube, flat-panel display, or a printed photograph), or the response for different types of sounds combine hard physical or chemical measurements with the subjective and varying responses of, say, the human eye or ear. However, these are precisely the quantities that a consumer uses to judge textiles, combinations of colored products, or the relative sound reproduction of music systems. They are therefore also the selling points of the marketer and innovator. How can the consumer choose and differentiate? How can they compare different claims? Semisubjective measurements such as these are moving away from the carefully controlled conditions of the laboratory into the high street and are presenting exciting new challenges.

NMIs are already becoming familiar with the needs of their users for color or acoustical measurement services, which require a degree of modeling of the user response and differing reactions depending on environmental conditions such as ambient lighting or noise background. The fascination of this area is that it combines objective metrology with physiological measurements and the inherent variability of the human eye or ear, or the ways in which our brains process optical or auditory stimuli.

Real-Time In-Process Measurements

The industries of today and tomorrow are starting to erode one of the century-old metrology practices within which the user must bring their own instruments and standards to the NMI for calibration. Some of this relates to the optimization of industrial processes, where far more accurate, real-time, in-process measurements are made. The economics of huge production processes demand just-in-time manufacture, active data management, and sophisticated process modeling. By reliably identifying where subelements of a process are behaving poorly, plant engineers can take rapid remedial action and so identify trouble spots quickly. However, actual real-time systems measurements are difficult, and it is only recently that some NMIs have begun to address the concept of an industrial measurement system. New business areas such as this will require NMIs to work differently, if for no other reason than because their customers work differently and they need to meet the customers' requirements. Remote telemetry, data fusion, and new sensor techniques are becoming linked with process modeling, numerical algorithms, and approximations so that accurate measurement can be put, precisely, at the point of measurement. These users are already adopting the systems approach, and some NMIs are starting to respond to this challenge.

Quantum-Based Standards

There is a trend towards quantum-based standards in industry, as already highlighted in this chapter. This is a result of the work of innovative instrument companies which now produce, for example, stabilized lasers, Josephson-junction voltage standards, and atomic clocks for the mass market. The availability of such highly accurate standards in industry is itself testimony to companies' relentless quest for improved product performance and quality. However, without care and experience, it is all too easy to get the wrong answer. Users are advised to undertake comparisons and to cooperate closely with their NMIs to make sure that these instruments are operated with all the proper checks and with attention to best practice so that they may, reliably, bring increased accuracy closer to the end user [2.4]. Industry presses NMIs – rightly so – for better performance, and in some areas of real practical need, NMI measurement capabilities are still rather close to what industry requires. It is, perhaps, in these highly competitive and market-driven areas that the key comparisons and statements of equivalence that are part of the CIPM MRA will prove their worth. Companies specify the performance of their products carefully in these highly competitive markets, and any significant differences in the way in which NMIs realize the SI units and quantities will have a direct bearing on competitiveness, market share, and profitability.

2.8.2 Chemistry, Pharmacy, and Medicine

Chemistry, biosciences, and pharmaceuticals are, for many of us, the new metrology. We are used to the practices of physical and engineering metrology, and so the new technologies are challenging our understanding of familiar concepts such as traceability, uncertainty, and primary standards. Much depends here on the interaction between a particular chemical or species and the solution or matrix in which it is to be found, as well as the processes or methods used to make the measurement. The concept of reference materials (RMs) (Chap. 3) is well developed in the chemical field, and the Consultative Committee for Quantity of Matter Metrology in Chemistry (CCQM) has embarked on a series of RM and other comparisons to assess the state of the art in chemical and biological measurements.

Through the CIPM MRA, the international metrology community has started to address the needs and concerns of regulators and legislators. This has already brought the Metre Convention into the areas of laboratory medicine and food [genetically modified organisms

(GMOs), pesticides, and trace elements]. The BIPM is only beginning to tackle and respond to the world of medicine and pharmacy and has created a partnership with the International Federation of Clinical Chemistry (IFCC) and ILAC to address these needs in a Joint Committee for Traceability in Laboratory Medicine (JCTLM). This is directed initially at a database of reference materials which meet certain common criteria of performance. Recognition of the data in the JCTLM database will, in particular, help demonstrate compliance of the products of the in vitro diagnostic industry with the requirements of a recent directive [2.5] of the European Union. The BIPM has signed a memorandum of understanding with the World Health Organization to help pursue these matters at an international level.

2.8.3 Environment, Public Services, and Infrastructures

In the area of the environment, our current knowledge of the complex interactions of weather, sea currents, and the various layers of our atmosphere is still not capable of a full explanation of environmental issues. The metrologist is, however, beginning to make a recognized contribution by insisting that slow or small changes, particularly in large quantities, should be measured traceably and against the unchanging reference stan-

dards offered through the units and quantities of the SI system. Similar inroads are being made into the space community where, for example, international and national space agencies are starting to appreciate that solar radiance measurements can be unreliable unless related to absolute measurements. We await the satellite launch of a cryogenic radiometer, which would do much to validate and monitor long-term trends in solar physics. The relevance of these activities is far from esoteric. Governments spend huge sums of money in their efforts to tackle environmental issues, and it is only by putting measurements on a sound basis that we can begin to make sure that these investments are justified and are making a real difference in the long term.

Traceable measurements are beginning to be needed as governments introduce competition into the supply of public services and infrastructures. Where utilities, such as electricity or gas, are offered by several different companies, the precise moment at which a consumer changes supplier can have large financial consequences. Traceable timing services are now used regularly in these industries as well as in stock exchanges and the growing world of e-commerce. Global trading is underpinned by globally accepted metrology, often without users appreciating the sophistication, reliability, and acceptability of the systems in which they place such implicit trust.

2.9 The SI System and New Science

Science never stands still, and there are a number of trends already evident which may have a direct bearing on the definitions of the SI itself. Much of this is linked with progress in measuring the fundamental constants, their coherence with each other and the SI, and the continued belief that they are time and space invariant. Figure 2.6 shows the current relationships of the basic SI units defined in Sect. 1.2.3 (Table 1.3).

This figure presents some of the links between the base units of the SI (shown in circles) and the fundamental physical and atomic constants (shown in boxes). It is intended to show that the base units of the SI are linked to measurable quantities through the unchanging and universal constants of physics.

The base units of the International System of Units defined in Table 1.3 are

A = ampere
K = Kelvin

s = second
m = meter
kg = kilogram
cd = candela
mol = mole

The surrounding boxes, lines, and uncertainties represent measurable quantities.

The numbers marked next to the base units are estimates of the standard uncertainties of their best practical realizations.

The fundamental and atomic constants shown are

R_K = von Klitzing constant
K_J = Josephson constant
R_{K-90} and K_{J-90} : the conventional values of these constants, introduced on 1 January 1990
N_A = Avogadro constant
F = Faraday constant

G	= Newtonian constant of gravitation
m_{12C}	= mass of ^{12}C
m_e	= electron mass
R_∞	= Rydberg constant
h	= Planck constant
c	= speed of light
μ_0	= magnetic constant
α	= fine-structure constant
R	= molar gas constant
k_B	= Boltzmann constant
e	= elementary charge

The numbers next to the fundamental and atomic constants represent the relative standard uncertainties of our knowledge of these constants (from the 2002 CODATA adjustment).

The grey boxes reflect the unknown long-term stability of the kilogram artifact and its consequent effects on the practical realization of the definitions of the ampere, mole, and candela.

Based on the fundamental SI units, derived physical quantities have been defined which can be expressed in

terms of the fundamental SI units, for example, the unit of force = mass \times acceleration is newton = mkg/s^2 , and the unit of work = force \times distance is joule = $\text{m}^2\text{kg/s}^2$. A compilation of the units which are used today in science and technology is given in the *Springer Handbook of Condensed Matter and Materials Data* [2.6].

There is considerable discussion at the moment [2.7] on the issue of redefinitions of a number of these base units. The driver is a wish to replace the kilogram artifact-based definition by one which assigns a fixed value to a fundamental constant, thereby following the precedent of the redefinition of the meter in 1983 which set the speed of light at 299 792 458 m/s. There are two possible approaches – the so-called watt balance experiment, which essentially is a measure of the Planck constant, and secondly a measurement of the Avogadro number, which can be linked to the Planck constant. For a detailed review of the two approaches and the scientific background see [2.8]. The two approaches, however, produce discrepant results, and there is insufficient convergence at the moment to enable a redefinition with uncertainty of a few parts in 10^8 . This uncertainty has been specified by the Consultative Committee for Mass

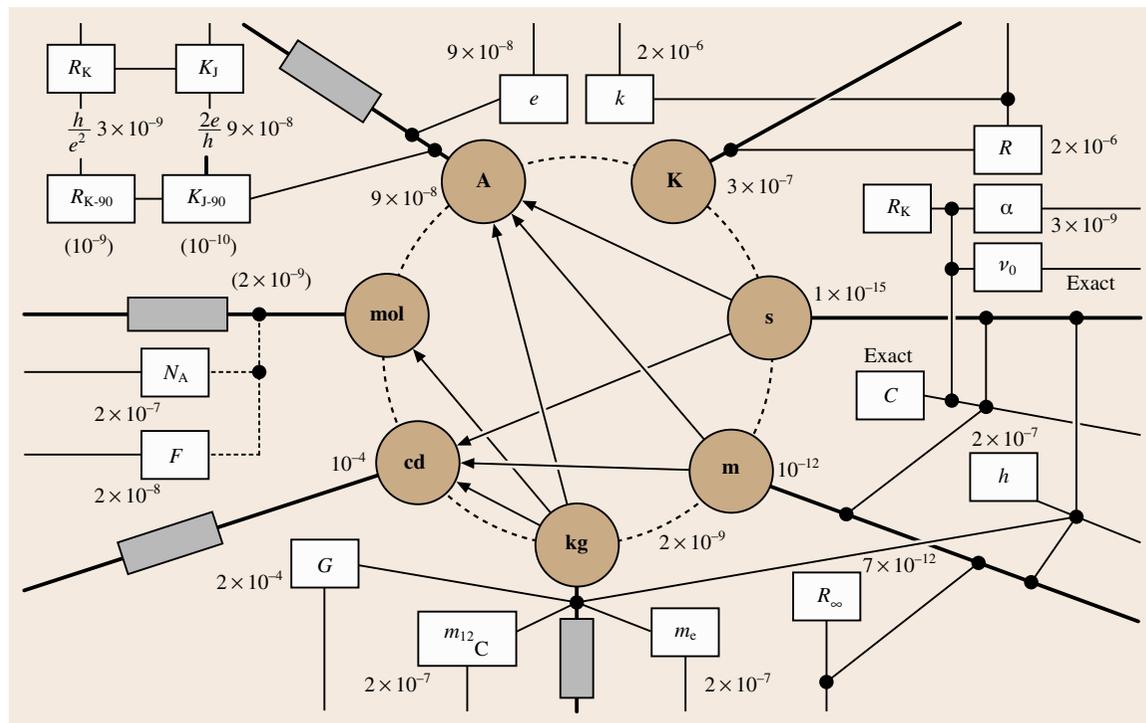


Fig. 2.6 Uncertainties of the fundamental constants (CODATA 2002, realization of the SI base units, present best estimates)

so as to provide for continuity of the uncertainty required and for minimum disturbance to the downstream world of practical mass measurements. If the Planck constant can be fixed, then it turns out that the electrical units can be related directly to the **SI** rather than based on the conventional values ascribed to the Josephson and von Klitzing constants by the **CIPM** in 1988 and universally referred to as R_{K-90} and J_{K-90} at the time of their implementation in 1990. A number of experiments are also underway to make improved measurements of the Boltzmann constant [2.9], which would be used to redefine the Kelvin, and finally a fixed value of the Planck constant with its connection to the Avogadro number would permit a redefinition of the mole.

This is, at the time of writing, a rapidly moving field, and any summary of the experimental situation would be rapidly overtaken by events. The reader is referred to the **BIPM** website (www.bipm.org), through which access to the deliberations of the relevant committees are available. However, the basic policy is now rather clear, and current thinking is that, within a few years and when the discrepancies are resolved, the General Conference on Weights and Measures will be asked to decide that four base units be redefined using

- a definition of the kilogram based on the Planck constant h ,
- a definition of the ampere based on a fixed value of the elementary charge e ,

- a definition of the Kelvin based on the value of the Boltzmann constant k_B , and
- a definition of the mole based on a fixed value of the Avogadro constant N_A .

The values ascribed to the fundamental constants would be those set by the CODATA group [2.10].

As a result, the *new SI* would be as shown diagrammatically in Fig. 2.7.

However, definitions are just what the word says – definitions – and there is a parallel effort to produce the *recipies* (called *mises en pratique*, following the precedent set at the time of redefining the meter) which enable their practical realizations worldwide.

When will all this take place? History will be the judge, but given current progress, the *new SI* could be in place by 2015. However, that will not be all.

New femtosecond laser techniques and the high performance of ion- or atom-trap standards may soon enable us to define the second in terms of optical transitions, rather than the current microwave-based one.

These are exciting times for metrologists. Trends in measurement are taking us into new regimes, and chemistry is introducing us to the world of reference materials and processes and to application areas which would have amazed our predecessors. What is, however, at face value surprising is the ability of a 130-year-old organization to respond flexibly and confidently to these

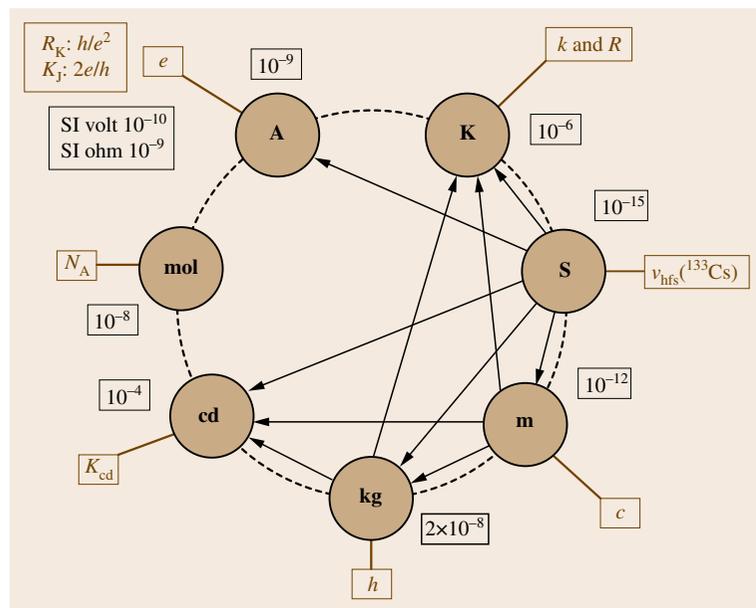


Fig. 2.7 Relations between the base units and fundamental constants together with the uncertainty associated with each link

changes. It is a tribute to our forefathers that the basis they set up for meter bars and platinum kilograms still applies to GMOs and measurement of cholesterol

in blood. Metrology is, truly, one of the oldest sciences and is one which continues to meet the changing needs of a changing world.

References

- 2.1 Bureau international des poids et mesures (BIPM): *The international system of units (SI)*, 7th edn. (BIPM, Sèvres 1998), see http://www.bipm.org/en/si/si_brochure
- 2.2 International Organization for Standardization (ISO): *International vocabulary of basic and general terms in metrology* (BIPM/IEC/IFCC/ISO/IUPAC/IUPAP/OIML, Genf 1993)
- 2.3 International Organization for Standardization (ISO): *Guide to the expression of uncertainty in measurement* (ISO, Genf 1993)
- 2.4 A.J. Wallard, T.J. Quinn: Intrinsic standards – Are they really what they claim?, *Cal Lab Mag.* **6**(6) (1999)
- 2.5 Directive 98/79/EC: See <http://www.bipm.org/en/committees/jc/jctlm>
- 2.6 W. Martienssen, H. Warlimont (Eds.): *Springer Handbook of Condensed Matter and Materials Data* (Springer, Berlin, Heidelberg 2005)
- 2.7 I.M. Mills, P.J. Mohr, T.J. Quinn, B.N. Taylor, E.R. Williams: Redefinition of the kilogram: A decision whose time has come, *Metrologia* **42**(2), 71–80 (2005)
- 2.8 http://www.sim-metrologia.org.br/docs/revista_SIM_2006.pdf
- 2.9 <http://www.bipm.org/wg/AllowedDocuments.jsp?wg=TG-SI>
- 2.10 <http://www.codata.org/index.html>

Quality in Me

3. Quality in Measurement and Testing

Technology and today's global economy depend on reliable measurements and tests that are accepted internationally. As has been explained in Chap. 1, metrology can be considered in categories with different levels of complexity and accuracy.

- Scientific metrology deals with the organization and development of measurement standards and with their maintenance.
- Industrial metrology has to ensure the adequate functioning of measurement instruments used in industry as well as in production and testing processes.
- Legal metrology is concerned with measurements that influence the transparency of economic transactions, health, and safety.

All scientific, industrial, and legal metrological tasks need appropriate *quality* methodologies, which are compiled in this chapter.

3.1	Sampling	40	3.4	Uncertainty and Accuracy of Measurement and Testing	68
	3.1.1 Quality of Sampling	40	3.4.1	General Principles	68
	3.1.2 Judging Whether Strategies of Measurement and Sampling Are Appropriate	42	3.4.2	Practical Example: Accuracy Classes of Measuring Instruments	69
	3.1.3 Options for the Design of Sampling	43	3.4.3	Multiple Measurement Uncertainty Components	71
3.2	Traceability of Measurements	45	3.4.4	Typical Measurement Uncertainty Sources	72
	3.2.1 Introduction	45	3.4.5	Random and Systematic Effects.....	73
	3.2.2 Terminology	46	3.4.6	Parameters Relating to Measurement Uncertainty: Accuracy, Trueness, and Precision ..	73
	3.2.3 Traceability of Measurement Results to SI Units.....	46	3.4.7	Uncertainty Evaluation: Interlaboratory and Intralaboratory Approaches	75
	3.2.4 Calibration of Measuring and Testing Devices	48	3.5	Validation	78
	3.2.5 The Increasing Importance of Metrological Traceability.....	49	3.5.1	Definition and Purpose of Validation	78
3.3	Statistical Evaluation of Results	50	3.5.2	Validation, Uncertainty of Measurement, Traceability, and Comparability.....	79
	3.3.1 Fundamental Concepts	50	3.5.3	Practice of Validation.....	81
	3.3.2 Calculations and Software	53	3.6	Interlaboratory Comparisons and Proficiency Testing	87
	3.3.3 Statistical Methods	54	3.6.1	The Benefit of Participation in PTs ..	88
	3.3.4 Statistics for Quality Control.....	66	3.6.2	Selection of Providers and Sources of Information	88
			3.6.3	Evaluation of the Results	92
			3.6.4	Influence of Test Methods Used.....	93
			3.6.5	Setting Criteria	94
			3.6.6	Trends	94
			3.6.7	What Can Cause Unsatisfactory Performance in a PT or ILC?.....	95
			3.6.8	Investigation of Unsatisfactory Performance	95
			3.6.9	Corrective Actions.....	96
			3.6.10	Conclusions	97
			3.7	Reference Materials	97
			3.7.1	Introduction and Definitions	97
			3.7.2	Classification	98
			3.7.3	Sources of Information	99
			3.7.4	Production and Distribution	100
			3.7.5	Selection and Use.....	101

3.7.6	Activities of International Organizations	104	3.9	Laboratory Accreditation and Peer Assessment	126
3.7.7	The Development of RM Activities and Application Examples	105	3.9.1	Accreditation of Conformity Assessment Bodies ...	126
3.7.8	Reference Materials for Mechanical Testing, General Aspects.....	107	3.9.2	Measurement Competence: Assessment and Confirmation	127
3.7.9	Reference Materials for Hardness Testing.....	109	3.9.3	Peer Assessment Schemes	130
3.7.10	Reference Materials for Impact Testing	110	3.9.4	Certification or Registration of Laboratories	130
3.7.11	Reference Materials for Tensile Testing	114	3.10	International Standards and Global Trade	130
3.8	Reference Procedures	116	3.10.1	International Standards and International Trade: The Example of Europe	131
3.8.1	Framework: Traceability and Reference Values ..	116	3.10.2	Conformity Assessment	132
3.8.2	Terminology: Concepts and Definitions	118	3.11	Human Aspects in a Laboratory	134
3.8.3	Requirements: Measurement Uncertainty, Traceability, and Acceptance	119	3.11.1	Processes to Enhance Competence – Understanding Processes.....	134
3.8.4	Applications for Reference and Routine Laboratories.....	121	3.11.2	The Principle of Controlled Input and Output.....	135
3.8.5	Presentation: Template for Reference Procedures.	123	3.11.3	The Five Major Elements for Consideration in a Laboratory ...	136
3.8.6	International Networks: CIPM and VAMAS	123	3.11.4	Internal Audits.....	136
3.8.7	Related Terms and Definitions	126	3.11.5	Conflicts.....	137
			3.11.6	Conclusions	137
			3.12	Further Reading: Books and Guides	138
			References		138

3.1 Sampling

Sampling is arguably the most important part of the measurement process. It is usually the case that it is impossible to measure the required quantity, such as concentration, in an entire batch of material. The taking of a sample is therefore the essential first step of nearly all measurements. However, it is commonly agreed that the quality of a measurement can be no better than the quality of the sampling upon which it is based. It follows that the highest level of care and attention paid to the instrumental measurements is ineffectual, if the original sample is of poor quality.

3.1.1 Quality of Sampling

The *traditional approach* to ensuring the quality of sampling is procedural rather than empirical. It relies initially on the selection of a correct sampling protocol for the particular material to be sampled under a particular circumstance. For example the material may be

copper metal, and the circumstance could be manufacturers' quality control prior to sale. In general, such a protocol may be specified by a regulatory body, or recommended in an international standard or by a trade organization. The second step is to train the personnel who are to take the samples (i.e., the samplers) in the correct application of the protocol. No sampling protocol can be completely unambiguous in its wording, so uniformity of interpretation relies on the samplers being educated, not just in how to interpret the words, but also in an appreciation of the rationale behind the protocol and how it can be adapted to the changing circumstances that will arise in the real world, without invalidating the protocol. This step is clearly related to the management of sampling by organizations, which is often separated from the management of the instrumental measurements, even though they are both inextricably linked to the overall quality of the measurement. The fundamental basis of the traditional approach

to assuring sampling quality is to assume that the correct application of a correct sampling protocol will give a representative sample, by definition.

An *alternative approach* to assuring sampling quality is to estimate the quality of sampling empirically. This is analogous to the approach that is routinely taken to instrumental measurement, where as well as specifying a protocol, there is an initial validation and ongoing quality control to monitor the quality of the measurements actually achieved. The key parameter of quality for instrumental measurements is now widely recognized to be the uncertainty of each measurement. This concept will be discussed in detail later (Sect. 3.4), but informally this uncertainty of measurement can be defined as the range within which the true value lies, for the quantity subject to measurement, with a stated level of probability. If the quantity subject to measurement (the measurand) is defined in terms of the batch of material (the sampling target), rather than merely in the sample delivered to the laboratory, then measurement uncertainty includes that arising from primary sampling. Given that sampling is the first step in the measurement process, then the uncertainty of the measurement will also arise in this first step, as well as in all of the other steps, such as the sampling preparation and the instrumental determination.

The key measure of sampling quality is therefore this sampling uncertainty, which includes contributions not just from the random errors often associated with sampling variance [3.1] but also from any systematic errors that have been introduced by sampling bias. Rather than assuming the bias is zero when the protocol is correct, it is more prudent to aim to include any bias in the estimate of sampling uncertainty. Such bias may often be unsuspected, and arise from a marginally incorrect application of a nominally correct protocol. This is equivalent to abandoning the assumption that samples are representative, but replacing it with a measurement result that has an associated estimate of uncertainty which includes errors arising from the sampling process.

Selection of the most appropriate sampling protocol is still a crucial issue in this alternative approach. It is possible, however, to select and monitor the appropriateness of a sampling protocol, by knowing the uncertainty of measurement that it generates. A judgement can then be made on the fitness for purpose (FFP) of the measurements, and hence the various components of the measurement process including the sampling, by comparing the uncertainty against the target value indicated by the FFP criterion. Two such FFP criteria are discussed below.

Two approaches have been proposed for the estimation of uncertainty from sampling [3.2]. The first or *bottom-up* approach requires the identification of all of the individual components of the uncertainty, the separate estimation of the contribution that each component makes, and then summation across all of the components [3.3]. Initial feasibility studies suggest that the use of sampling theory to predict all of the components will be impractical for all but a few sampling systems, where the material is particulate in nature and the system conforms to a model in which the particle size/shape and analyte concentration are simple, constant, and homogeneously distributed. One recent application successfully mixes theoretical and empirical estimation techniques [3.4]. The second, more practical and pragmatic approach is entirely empirical, and has been called *top-down* estimation of uncertainty [3.5].

Four methods have been described for the empirical estimation of uncertainty of measurement, including that from primary sampling [3.6]. These methods can be applied to any sampling protocol for the sampling of any medium for any quantity, if the general principles are followed. The simplest of these methods (#1) is called the *duplicate method*. At its simplest, a small proportion of the measurements are made in duplicate. This is not just a duplicate analysis (i. e., determination of the quantity), made on one sample, but made on a fresh primary sample, from the same sampling target as the original sample, using a fresh interpretation of the same sampling protocol (Fig. 3.1a). The ambiguities in the protocol, and the heterogeneity of the material, are therefore reflected in the difference between the duplicate measurements (and samples). Only 10% ($n \geq 8$) of the samples need to be duplicated to give a sufficiently reliable estimate of the overall uncertainty [3.7]. If the separate sources of the uncertainty need to be quantified, then extra duplication can be inserted into the experimental design, either in the determination of quantity (Fig. 3.1b) or in other steps, such as the physical preparation of the sample (Fig. 3.1d). This duplication can either be on just one sample duplicate (in an *unbalanced* design, Fig. 3.1b), or on both of the samples duplicated (in a *balanced* design, Fig. 3.1c).

The uncertainty of the measurement, and its components if required, can be estimated using the statistical technique called analysis of variance (ANOVA). The frequency distribution of measurements, such as analyte concentration, often deviate from the normal distribution that is assumed by classical ANOVA. Because of this, special procedures are required to accommodate outlying values, such as robust ANOVA [3.8]. This method

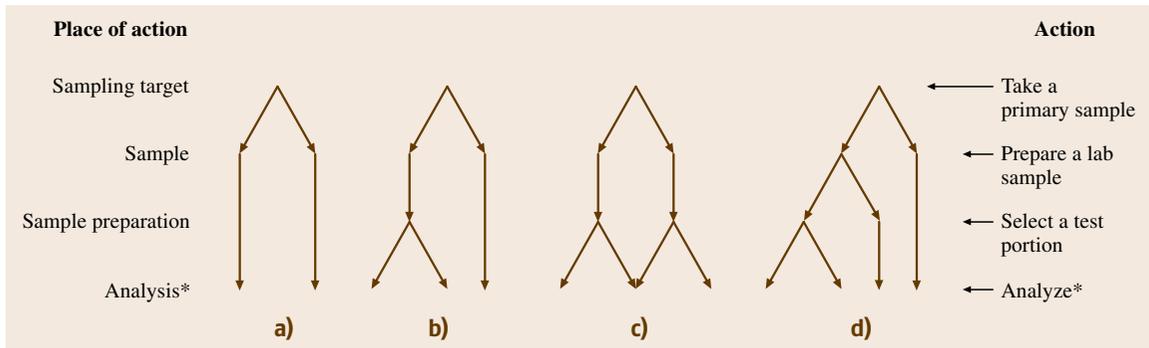


Fig. 3.1a–d Experimental designs for the estimation of measurement uncertainty by the duplicate method. The simplest and cheapest option (a) has single analyses on duplicate samples taken on around 10% ($n \geq 8$) of the sampling targets, and only provides an estimate of the random component of the overall measurement uncertainty. If the contribution from the analytical determination is required separately from that from the sampling, duplication of analysis is required on either one (b) or both (c) of the sample duplicates. If the contribution from the physical sample preparation is required to be separated from the sampling, as well as from that from the analysis, then duplicate preparations also have to be made (d). (*Analysis and Analyze can more generally be described as the determination of the measurand)

has successfully been applied to the estimation of uncertainty for measurements on soils, groundwater, animal feed, and food materials [3.2]. Its weakness is that it ignores the contribution of systematic errors (from sampling or analytical determination) to the measurement uncertainty. Estimates of analytical bias, made with certified reference materials, can be added to estimates from this method. Systematic errors caused by a particular sampling protocol can be detected by use of a different method (#2) in which different sampling protocols are applied by a single sampler. Systematic errors caused by the sampler can also be incorporated into the estimate of measurement uncertainty by the use of a more elaborate method (#3) in which several samplers apply the same protocol. This is equivalent to holding a *collaborative trial in sampling (CTS)*. The most reliable estimate of measurement uncertainty caused by sampling uses the most expensive method (#4), in which several samplers each apply whichever protocol they consider most appropriate for the stated objective. This incorporates possible systematic errors from the samplers and the measurement protocols, together with all of the random errors. It is in effect a sampling proficiency test (SPT), if the number of samplers is at least eight [3.6].

Evidence from applications of these four empirical methods suggests that small-scale heterogeneity is often the main factor limiting the uncertainty. In this case, methods that concentrate on repeatability, even with just one sampler and one protocol as in the duplicate method (#1), are good enough to give an acceptable approximation of the sampling uncertainty. Proficiency

test measurements have also been used in *top-down* estimation of uncertainty of analytical measurements [3.9]. They do have the added advantage that the participants are scored for the proximity of their measurement value to the true value of the quantity subject to measurement. This true value can be estimated either by consensus of the measured values, or by artificial spiking with a known quantity of analyte [3.10]. The score from such SPTs could also be used for both ongoing assessment and accreditation of samplers [3.11]. These are all new approaches that can be applied to improving the quality of sampling that is actually achieved.

3.1.2 Judging Whether Strategies of Measurement and Sampling Are Appropriate

Once methods are in place for the estimation of uncertainty, the selection and implementation of a correct protocol become less crucial. Nevertheless an appropriate protocol is essential to achieve fitness for purpose. The FFP criterion may however vary, depending on the circumstances. There are cases for example where a relative expanded uncertainty of 80% of the measured value can be shown to be fit for certain purposes. One example is using in situ measurements of lead concentration to identify any area requiring remediation in a contaminated land investigation. The contrast between the concentration in the contaminated and in the uncontaminated areas can be several orders of magnitude, and so uncertainty within one order (i.e., 80%) does not

result in errors in classification of the land. A similar situation applies when using laser-ablation inductively coupled plasma for the determination of silver to differentiate between particles of anode copper from widely different sources. The Ag concentration can differ by several orders of magnitude, so again a large measurement uncertainty (e.g., 70%) can be acceptable. One mathematical way of expressing this FFP criterion is that the measurement uncertainty should not contribute more than 20% to the total variance over samples from a set of similar targets [3.8]. A second FFP criterion also includes financial considerations, and aims to set an optimal level of uncertainty that minimizes financial loss. This loss arises not just from the cost of the sampling and the determination, but also from the financial losses that may arise from incorrect decisions caused by the uncertainty [3.12]. The approach has been successfully applied to the sampling of both contaminated soil [3.13] and food materials [3.14].

3.1.3 Options for the Design of Sampling

There are three basic approaches to the design/selection of a sampling protocol for any quantity (measurand) in any material. The first option is to select a previously specified protocol. These exist for most of the material/quantity combinations considered in Chap. 4 of this handbook. This approach is favored by regulators, who expect that the specification and application of a standard protocol will automatically deliver comparability of results between samplers. It is also used as a defense in legal cases to support the contention that measurements will be reliable if a standard protocol has been applied. The rationale of a standard protocol is to specify the procedure to the point where the sampler needs to make no subjective judgements. In this case the sampler would appear not to require any grasp of the rationale behind the design of the protocol, but merely the ability to implement the instructions given. However, experimental video monitoring of samplers implementing specified protocols suggests that individual samplers often do extemporize, especially when events occur that were unforeseen or unconsidered by the writers of the protocols. This would suggest that samplers therefore need to appreciate the rationale behind the design, in order to make appropriate decisions on implementing the protocol. This relates to the general requirement for improved training and motivation of samplers discussed below.

The second option is to use a theoretical model to design the required sampling protocol. Sampling

theory has produced a series of increasingly complex theoretical models, recently reviewed [3.15], that are usually aimed at predicting the sampling mass required to produce a given level of variance in the required measurement result. All such models depend on several assumptions about the system that is being modeled. The model of Gy [3.1], for example, assumes that the material is particulate, that the particles in the batch can be classified according to volume and type of material, and that the analyte concentration in a contaminated particle and its density do not vary between particles. It was also assumed that the volume of each particle in the batch is given by a constant factor multiplied by the cube of the particle diameter. The models also all require large amounts of information about the system, such as particle diameters, shape factors, size range, liberation, and composition. The cost of obtaining all of this information can be very high, but the model also assumes that these parameters will not vary in space or time. These assumptions may not be justified for many systems in which the material to be sampled is highly complex, heterogeneous, and variable. This limits the real applicability of this approach for many materials. These models do have a more generally useful role, however, in facilitating the prediction of how uncertainty from sampling can be changed, if required, as discussed below.

The third option for designing a sampling protocol is to adapt an existing method in the light of site-specific information, and monitor its effectiveness empirically. There are several factors that require consideration in this adaptation.

Clearly identifying the *objective of the sampling* is the key factor that helps in the design of the most appropriate sampling protocol. For example, it may be that the acceptance of a material is based upon the best estimate of the *mean* concentration of some analyte in a batch. Alternatively, it may be the *maximum* concentration, within some specified mass, that is the basis for acceptance or rejection. Protocols that aim at low uncertainty in estimation of the mean value are often inappropriate for reliable detection of the maximum value.

A *desk-based review* of all of the relevant information about the sampling target, and findings from similar targets, can make the protocol design much more cost effective. For example, the history of a contaminated land site can suggest the most likely contaminants and their probable spatial distribution within the site. This information can justify using *judgemental sampling* in which the highest sampling density is concentrated in

the area of highest predicted probability. This approach does however, have the weakness that it may be self-fulfilling, by missing contamination in areas that were unsuspected.

The actual *mode of sampling* varies greatly therefore, depending not just on the physical nature of the materials, but also on the expected heterogeneity in both the spatial and temporal dimension. Some protocols are designed to be random (or nonjudgemental) in their selection of samples, which in theory creates the least bias in the characterization of the measurand. There are various different options for the design of random sampling, such as stratified random sampling, where the target is subdivided into regular units before the exact location of the sampling is determined using randomly selected coordinates. In a situation where judgemental sampling is employed, as described above, the objective is not to get a representative picture of the sampling target. Another example would be in an investigation of the cause of defects in a metallurgical process, where it may be better to select items within a batch by their aberrant visual appearance, or contaminant concentration, rather than at random.

There may also be a question of the most appropriate *medium to sample*. The answer may seem obvious, but consider the objective of detecting which of several freight containers holds nuts that are contaminated with mycotoxins. Rather than sampling the nuts themselves, it may be much more cost effective to sample the atmosphere in each container for the spores released by the fungi that make the mycotoxin. Similarly in contaminated land investigation, if the objective is to assess potential exposure of humans to cadmium at an allotment site, it may be most effective to sample the vegetables that take up the cadmium rather than the soil.

The *specification of the sampling target* needs to be clear. Is it a whole batch, or a whole site of soil, or just the top 1 m of the soil? This relates to the objective of the sampling, but also to the site-specific information (e.g., there is bedrock at 0.5 m) and logistical constraints.

The next key question to address is the *number of samples* required (n). This may be specified in an accepted sampling protocol, but should really depend on the objective of the investigation. Cost-benefit analysis can be applied to this question, especially if the objective is the mean concentration at a specified confidence interval. In that case, and assuming a normal distribution of the variable, the Student t -distribution can be used to calculate the required value of n . A closely related question is whether *composite samples* should be

taken, and if so, what is the required number of increments (i). This approach can be used to reduce the uncertainty of measurement caused by the sampling. According to the theory of Gy, taking an i -fold composite sample should reduce the main source of the uncertainty by \sqrt{i} , compared with the uncertainty for a single sample with the same mass as one of the increments. Not only do the increments increase the sample mass, but they also improve the sample's ability to represent the sampling target. If, however, the objective is to identify maximum rather than mean values, then a different approach is needed for calculating the number of samples required. This has been addressed for contaminated land by calculating the probability of hitting an idealized hot-spot [3.16].

The *quantity of sample* to be taken (e.g., mass or volume) is another closely related consideration in the design of a specified protocol. The mass may be specified by existing practise and regulation, or calculated from sampling theory such as that of Gy. Although the calculation of the mass from first principles is problematic for many types of sample, as already discussed, the theory is useful in calculating the factor by which to change the sample mass to achieve a specified target for uncertainty. If the mass of the sample is increased by some factor, then the sampling variance should reduce by the same factor, as discussed above for increments. The mass required for measurement is often smaller than that required to give an acceptable degree of representativeness (and uncertainty). In this case, a larger sample must be taken initially and then reduced in mass, without introducing bias. This *comminution of samples*, or reduction in grain size by grinding, is a common method for reducing the uncertainty introduced by this subsampling procedure. This can, however, have unwanted side-effects in changing the measurand. One example is the loss of certain analytes during the grinding, either by volatilization (e.g., mercury) or by decomposition (e.g., most organic compounds).

The *size of the particles* in the original sampling target that should constitute the sample needs consideration. Traditional wisdom may suggest that a representative sample of the whole sampling target is required. However, sampling all particle sizes in the same proportions that they occur in the sampling target may not be possible. This could be due to limitations in the sampling equipment, which may exclude the largest particles (e.g., pebbles in soil samples). A representative sample may not even be desirable, as in the case where only the small particles in soil ($< 100 \mu\text{m}$) form the main route of human exposure to lead by hand-to-

mouth activity. The objectives of the investigation may require therefore that a specific size fraction be selected.

Contamination of samples is probable during many of these techniques of sampling processing. It is often easily done, irreversible in its effect, and hard to detect. It may arise from other materials at the sampling site (e.g., topsoil contaminating subsoil) or from processing equipment (e.g., cadmium plating) or from the remains of previous samples left in the equipment. The traditional approach is to minimize the risk of contamination occurring by careful drafting of the protocol, but a more rigorous approach is to include additional procedures that can detect any contamination that has occurred (e.g., using a **SPT**).

Once a sample has been taken, the protocol needs to describe how to *preserve the sample*, without changing the quantity subject to measurement. For some measurands the quantity begins to change almost immediately after sampling (e.g., the redox potential of groundwater), and in situ measurement is the most reliable way of avoiding the change. For other measurands specific actions are required to prevent change. For example, acidification of water, after filtration, can prevent adsorption of many analyte ions onto the surfaces of a sample container.

The final, and perhaps most important factor to consider in designing a sampling protocol is the logistical

organization of the samples within the investigation. Attention to detail in the unique numbering and clear description of samples can avoid ambiguity and irreversible errors. This improves the quality of the investigation by reducing the risk of gross errors. Moreover, it is often essential for legal traceability to establish an unbroken chain of custody for every sample. This forms part of the broader quality assurance of the sampling procedure.

There is no such thing as either a perfect sample or a perfect measurement. It is better, therefore, to estimate the uncertainty of measurements from all sources, including the primary sampling. The uncertainty should not just be estimated in an initial method validation, but also monitored routinely for every batch using a sampling and analytical quality control scheme (**SAQCS**). This allows the investigator to judge whether each batch of measurements are **FFP**, rather than to assume that they are because some standard procedure was nominally adhered to. It also enables the investigator to propagate the uncertainty value through all subsequent calculations to allow the uncertainty on the interpretation of the measurements to be expressed. This approach allows for the imperfections in the measurement methods and the humans who implement them, and also for the heterogeneity of the real world.

3.2 Traceability of Measurements

3.2.1 Introduction

Clients of laboratories will expect that results are correct and comparable. It is further anticipated that complete results and values produced include an estimated uncertainty. A comparison between different results or between results achieved and given specifications can only be done correctly if the measurement uncertainty of the results is taken into account.

To achieve comparable results, the traceability of the measurement results to **SI** units through an unbroken chain of comparisons, all having stated uncertainties, is fundamental (Sect. 2.6 *Traceability of Measurements*). Among others, due to the strong request from the International Laboratory Accreditation Cooperation (**ILAC**) several years ago, the International Committee for Weights and Measures (**CIPM**), which is the governing board of the International Bureau of Weights and Measures (**BIPM**), has realized under the scope

of the Metre Convention the **CIPM** mutual recognition arrangement (**MRA**) on the mutual recognition of national measurement standards and of calibration and measurement certificates issued by the national metrology institutes, under the scope of the Metre Convention. Details of this **MRA** can be found in Chap. 2 *Metrology Principles and Organization* Sect. 2.7 or at <http://www1.bipm.org/en/convention/mra/>.

The range of national measurement standards and best measurement capabilities needed to support the calibration and testing infrastructure in an economy or region can normally be derived from the websites of the respective national metrology institute or from the website of the **BIPM**. Traceability to these national measurement standards through an unbroken chain of comparisons is an important means to achieve accuracy and comparability of measurement results.

Access to suitable national measurement standards may be more complicated in those economies where

the national measurement institute does not yet provide national measurement standards recognized under the **BIPM MRA**. It is further to be noted that an unbroken chain of comparisons to national standards in various fields such as the chemical and biological sciences is much more complex and often not available, as appropriate standards are lacking. The establishment of standards in these fields is still the subject of intense scientific and technical activities, and reference procedures and (certified) reference materials needed must still be defined. As of today, in these fields there are few reference materials that can be traced back to **SI** units available on the market. This means that other tools should also be applied to assure at least comparability of measurement results, such as, e.g., participation in suitable proficiency testing programs or the use of reference materials provided by reliable and competent reference material producers.

3.2.2 Terminology

According to the *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM 2008)* [3.17], the following definitions apply.

Primary Measurement Standard

Measurement standard established using a primary reference measurement procedure, or created as an artifact, chosen by convention.

International Measurement Standard

Measurement standard recognized by signatories to an international agreement and intended to serve worldwide.

National Measurement Standard, National Standard

Measurement standard recognized by national authority to serve in a state or economy as the basis for assigning quantity values to other measurement standards for the kind of quantity concerned.

Reference Measurement Standard, Reference Standard

Measurement standard designated for the calibration of other measurement standards for quantities of a given kind in a given organization or at a given location.

Working Standard

Measurement standard that is used routinely to calibrate or verify measuring instruments or measuring systems.

Note that a working standard is usually calibrated against a reference standard. Working standards may also at the same time be reference standards. This is particularly the case for working standards directly calibrated against the standards of a national standards laboratory.

3.2.3 Traceability of Measurement Results to SI Units

The formal definition of traceability is given in Chap. 2, Sect. 2.6 as: the property of a measurement result relating the result to a stated metrological reference through an unbroken chain of calibrations or comparisons, each contributing to the stated uncertainty. This chain is also called the traceability chain. It must, as defined, end at the respective primary standard.

The uncertainty of measurement for each step in the traceability chain must be calculated or estimated according to agreed methods and must be stated so that an overall uncertainty for the whole chain may be calculated or estimated. The calculation of uncertainty is officially given in the *Guide to the Expression of Uncertainty in Measurement (GUM)* [3.18]. The **ILAC** and regional organizations of accreditation bodies (see under peer and third-party assessment) provide application documents derived from the **GUM**, providing instructive examples. These documents are available on their websites.

Competent testing laboratories, e.g., those accredited by accreditation bodies that are members of the **ILAC MRA**, can demonstrate that calibration of equipment that makes a significant contribution to the uncertainty and hence the measurement results generated by that equipment are traceable to the international system of units (**SI** units) wherever this is technically possible.

In cases where traceability to the **SI** units is not (yet) possible, laboratories use other means to assure at least comparability of their results. Such means are, e.g., the use of certified reference materials, provided by a reliable and competent producer, or they assure at least comparability by participating in interlaboratory comparisons provided by a competent and reliable provider. See also Sects. 3.6 and 3.7 on *Interlaboratory Comparisons and Proficiency Testing* and *Reference Materials*, respectively.

The Traceability Chain

National Metrology Institutes. In most cases the national metrology institutes maintain the *national standards* that are the sources of traceability for the quantity

of interest. The national metrology institutes ensure the comparability of these standards through an international system of key comparisons, as explained in detail in Chap. 2, Sect. 2.7.

If a national metrology institute has an infrastructure to realize a given primary standard itself, this national standard is identical to or directly traceable to that *primary standard*. If the institute does not have such an infrastructure, it will ensure that its national standard is traceable to a primary standard maintained in another country's institute. Under <http://kcdb.bipm.org/AppendixC/default.asp>, the calibration and measurement capabilities (CMCs) declared by national metrology institutes are shown.

Calibration Laboratories. For *calibration laboratories* accredited according to the ISO/International Electrotechnical Commission (IEC) standard ISO/IEC 17025, accreditation is granted for specified calibrations with a defined calibration capability that can (but not necessarily must) be achieved with a specified measuring instrument, reference or working standard.

The *calibration capability* is defined as the smallest uncertainty of measurement that a laboratory can achieve within its scope of accreditation, when performing more or less routine calibrations of nearly ideal measurement standards intended to realize, conserve or reproduce a unit of that quantity or one or more of its values, or when performing more or less routine calibrations of nearly ideal measuring instruments designed for the measurement of that quantity.

Most of the accredited laboratories provide calibrations for customers (e.g., for organizations that do not have their own calibration facilities with a suitable measurement capability or for testing laboratories) on request. If the service of such an accredited calibration laboratory is taken into account, it must be assured that its scope of accreditation fits the needs of the customer. Accreditation bodies are obliged to provide a list of accredited laboratories with a detailed technical description of their scope of accreditation. <http://www.ilac.org/> provides a list of the accreditation bodies which are members of the ILAC MRA.

If a customer is using a nonaccredited calibration laboratory or if the scope of accreditation of a particular calibration laboratory does not fully cover a specific calibration required, the customer of that laboratory must ensure that

- the tractability chain as described above is maintained correctly,

- there is a concept to estimate the overall measurement uncertainty in place and applied correctly,
- the staff is thoroughly trained to perform the activities within their responsibilities,
- clear and valid procedures are available to perform the required calibrations,
- a system to deal with errors is applied, and the calibration operations include statistical process control such as, e.g., the use of control charts.

In-House Calibration Laboratories (Factory Calibration Laboratories). Frequently, calibration services are provided by in-house calibration laboratories which regularly calibrate the measuring and test equipment used in a company, e.g., in a production facility, against its reference standards that are traceable to an accredited calibration laboratory or a national metrology institute.

An in-house calibration system normally assures that all measuring and test equipment used within a company is calibrated regularly against working standards, calibrated by an accredited calibration laboratory. In-house calibrations must fit into the internal applications in such a way that the results obtained with the measuring and test equipment are accurate and reliable. This means that for in-house calibration the following elements should be considered as well.

- The uncertainty contribution of the in-house calibration should be known and taken into account if statements of compliance, e.g., internal criteria for measuring instruments, are made.
- The staff should be trained to perform the calibrations required correctly.
- Clear and valid procedures should be available also for in-house calibrations.
- A system to deal with errors should be applied (e.g., in the frame of an overall quality management system), and the calibration operations should include a statistical process control (e.g., the use of control charts).

To assure correct operation of the measuring and test equipment, a concept for the maintenance of that equipment should be in place. Aspects to be considered when establishing calibration intervals are given in Sect. 3.5.

The Hierarchy of Standards. The hierarchy of standards and a resulting metrological organizational structure for tracing measurement and test results within a company to national standards are shown in Fig. 3.2.

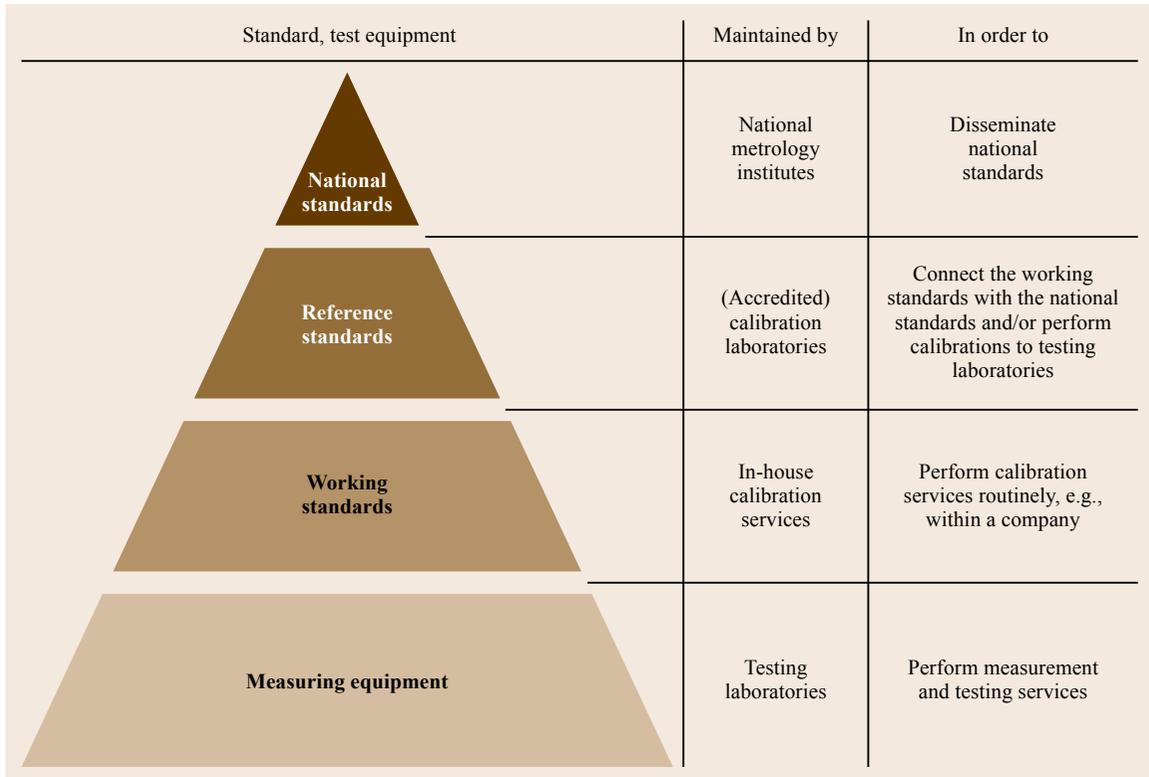


Fig. 3.2 The calibration hierarchy

Equipment used by testing and calibration laboratories that has a significant effect on the reliability and uncertainty of measurement should be calibrated using standards connected to the national standards with a known uncertainty.

Alternative Solutions

Accreditation bodies which are members of the [ILAC MRA](#) require accredited laboratories to ensure traceability of their calibration and test results. Accredited laboratories also know the contribution of the uncertainty derived through the traceability chain to their calibration and test results.

Where such traceability is not (yet) possible, laboratories should at least assure comparability of their results by alternative methods. This can be done either through the use of appropriate reference materials ([RM](#)) or by participating regularly in appropriate proficiency tests ([PT](#)) or interlaboratory comparisons. *Appropriate* means that the [RM](#) producers or the [PT](#) providers are competent or at least recognized in the respective sector.

3.2.4 Calibration of Measuring and Testing Devices

The [VIM 2008](#) gives the following definition for calibration:

Definition

Operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication.

The operation of calibration and its two steps is described in [Sect. 3.4.2](#) with an example from dimensional metrology ([Fig. 3.10](#)).

It is common and important that testing laboratories regularly maintain and control their testing instruments, measuring systems, and reference and working standards. Laboratories working according to the [ISO/IEC 17025](#) standard as well as manufactur-

ers working according to, e.g., the ISO 9001 series of standards maintain and calibrate their measuring instruments, and reference and working standards regularly according to well-defined procedures.

Clause 5.5.2 of the ISO/IEC 17025 standard requires that:

Calibration programmes shall be established for key quantities or values of the instruments where these properties have a significant effect on the results.

Whenever practicable, all equipment under the control of the laboratory and requiring calibration shall be labeled, coded, or otherwise identified to indicate the status of calibration, including the data when last calibrated and the date or expiration criteria when recalibration is due. (Clause 5.5.8)

Clause 7.6 of ISO 9001:2000 requires that:

Where necessary to ensure valid results, measuring equipment shall be calibrated or verified at specified intervals, or prior to use, against measurement standards traceable to international or national measurement standards.

In the frame of the calibration programs of their measuring instruments, and reference and working standards, laboratories will have to define the time that should be permitted between successive calibrations (recalibrations) of the used measurement instruments, and reference or working standards in order to

- confirm that there has not been any deviation of the measuring instrument that could introduce doubt about the results delivered in the elapsed period,
- assure that the difference between a reference value and the value obtained using a measuring instrument is within acceptable limits, also taking into account the uncertainties of both values,
- assure that the uncertainty that can be achieved with the measuring instrument is within expected limits.

A large number of factors can influence the time interval to be defined between calibrations and should be taken into account by the laboratory. The most important factors are usually

- the information provided by the manufacturer,
- the frequency of use and the conditions under which the instrument is used,

- the risk of the measuring instrument drifting out of the accepted tolerance,
- consequences which may arise from inaccurate measurements (e.g., failure costs in the production line or aspects of legal liability),
- the cost of necessary corrective actions in case of drifting away from the accepted tolerances,
- environmental conditions such as, e.g., climatic conditions, vibration, ionizing radiation, etc.,
- trend data obtained, e.g., from previous calibration records or the use of control charts,
- recorded history of maintenance and servicing,
- uncertainty of measurement required or declared by the laboratory.

These examples show the importance of establishing a concept for the maintenance of the testing instruments and measuring systems. In the frame of such a concept the definition of the calibration intervals is one important aspect to consider. To optimize the calibration intervals, available statistical results, e.g., from the use of control charts, from participation in interlaboratory comparisons or from reviewing own records should be used.

3.2.5 The Increasing Importance of Metrological Traceability

An increasing awareness of the need for metrological underpinning of measurements can be noticed at least in the past years. Several factors may be the reason for this process, including

- the importance of quality management systems,
- requirements by governments or trading partners for producers to establish certified quality management systems and for calibration and testing activities to be accredited,
- aspects of legal reliability.

In a lot of areas it is highly important that measurement results, e.g., produced by testing laboratories, can be compared with other results produced by other parties at another time and quite often using different methods. This can only be achieved if measurements are based on equivalent physical realizations of units. Traceability of results and reference values to primary standards is a fundamental issue in competent laboratory operation today.

3.3 Statistical Evaluation of Results

Statistics are used for a variety of purposes in measurement science, including mathematical modeling and prediction for calibration and method development, method validation, uncertainty estimation, quality control and assurance, and summarizing and presenting results. This section provides an introduction to the main statistical techniques applied in measurement science. A knowledge of the basic descriptive statistics (mean, median, standard deviation, variance, quantiles) is assumed.

3.3.1 Fundamental Concepts

Measurement Theory and Statistics

The traditional application of statistics to quantitative measurement follows a set of basic assumptions related to ordinary statistics

1. That a given measurand has a value – the value of the measurand – which is unknown and (in general) unknowable by the measurement scientist. This is generally assumed (for univariate quantitative measurements) to be a single value for the purpose of statistical treatment. In statistical standards, this is the *true value*.
2. That each measurement provides an *estimate* of the value of the measurand, formed from an *observation* or set of observations.
3. That an observation is the sum of the measurand value and an *error*.

Assumption 3 can be expressed as one of the simplest statistical models

$$x_i = \mu + e_i,$$

in which x_i is the i -th observation, μ is the measurand value, and e_i is the error in the particular observation.

The error itself is usually considered to be a sum of several contributions from different sources or with different behavior. The most common partition of error is into two parts: one which is constant for the duration of a set of experiments (the *systematic error*) and another, the *random error*, which is assumed to arise by random selection from some distribution. Other partitioning is possible; for example, collaborative study uses a statistical model based on a systematic contribution (method bias), a term which is constant for a particular laboratory (the laboratory component of bias) but randomly distributed among laboratories, and a residual error for each observation. Linear calibration

assumes that observations are the sum of a term that varies linearly and systematically with measurand value and a random term; least-squares regression is one way of characterizing the behavior of the systematic part of this model.

The importance of this approach is that, while the value of the measurand may be unknown, studying the distribution of the observations allows inferences to be drawn about the probable value of the measurand. Statistical theory describes and interrelates the behaviors of different distributions, and this provides quantitative tools for describing the probability of particular observations given certain assumptions. Inferences can be drawn about the value of the measurand by asking what range of measurand values could reasonably lead to the observations found. This provides a range of values that can reasonably be attributed to the measurand. Informed readers will note that this is the phrase used in the definition of uncertainty of measurement, which is discussed further below.

This philosophy forms the basis of many of the routine statistical methods applied in measurement, is well established with strong theoretical foundations, and has stood the test of time well. This chapter will accordingly rely heavily on the relevant concepts. It is, however, important to be aware that it has limitations. The basic assumption of a point value for the measurand may be inappropriate for some situations. The approach does not deal well with the accumulation of information from a variety of different sources. Perhaps most importantly, real-world data rarely follow theoretical distributions very closely, and it can be misleading to take inference too far, and particularly to infer very small probabilities or very high levels of confidence. Furthermore, other theoretical viewpoints can be taken and can provide different insights into, for example, the development of confidence in a value as data from different experiments are accumulated, and the treatment of estimates based on judgement instead of experiment.

Distributions

Figure 3.3 shows a typical measurement data set from a method validation exercise. The tabulated data shows a range of values. Plotting the data in histogram form shows that observations tend to cluster near the center of the data set. The histogram is one possible graphical representation of the *distribution* of the data.

If the experiment is repeated, a visibly different data distribution is usually observed. However, as the

number of observations in an experiment increases, the distribution becomes more consistent from experiment to experiment, tending towards some underlying form. This underlying form is sometimes called the *parent* distribution. In Fig. 3.3, the smooth curve is a plot of a possible parent distribution, in this case, a *normal distribution* with a mean and standard deviation estimated from the data.

There are several important features of the parent distribution shown in Fig. 3.3. First, it can be represented by a mathematical equation – a distribution function – with a relatively small number of parameters. For the normal distribution, the parameters are the mean and population standard deviation. Knowing that the parent distribution is normal, it is possible to summarize a large number of observations simply by giving the mean and standard deviation. This allows large sets of observations to be summarized in terms of the distribution type and the relevant parameters. Second, the distribution can be used predictively to make statements about the likelihood of further observations; in Fig. 3.3, for example, the curve indicates that observations in the region of 2750–2760 mg kg⁻¹ will occur only rarely. The distribution is accordingly important in both describing data and in drawing inferences from the data.

Distributions of Measurement Data. Measurement data can often be expected to follow a normal distribution, and in considering statistical tests for ordinary

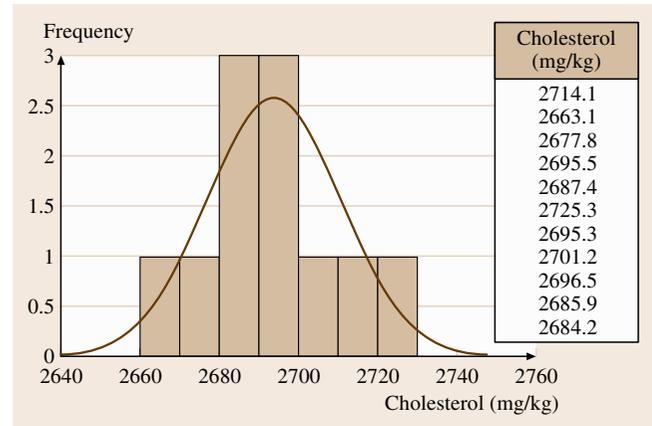


Fig. 3.3 Typical measurement data. Data from 11 replicate analyses of a certified reference material with a certified value of 2747±90 mg/kg cholesterol. The *curve* is a normal distribution with mean and standard deviation calculated from the data, with vertical scaling adjusted for comparability with the histogram

cases, this will be the assumed distribution. However, some other distributions are important in particular circumstances. Table 3.1 lists some common distributions, whose general shape is shown in Fig. 3.4. The most important features of each are

- The normal distribution is described by two independent parameters: the mean and standard deviation. The mean can take any value, and the standard

Table 3.1 Common distributions in measurement data

Distribution	Density function	Mean	Expected variance	Remarks
Normal	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	μ	σ^2	Arises naturally from the summation of many small random errors from any distribution
Lognormal	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)$	$\exp\left(\mu + \frac{\sigma^2}{2}\right)$	$\exp(2\mu + \sigma^2) [\exp(\sigma^2) - 1]$	Arises naturally from the product of many terms with random errors. Approximates to normal for small standard deviation
Poisson	$\lambda^x \exp(-\lambda)/x!$	λ	λ	Distribution of events occurring in an interval; important for radiation counting. Approximates to normality for large λ
Binomial	$\binom{n}{x} p^x (1-p)^{(n-x)}$	np	$np(1-p)$	Distribution of x , the number of successes in n trials with probability of success p . Common in counting at low to moderate levels, such as microbial counts; also relevant in situations dominated by particulate sampling
Contaminated normal	Various			Contaminated normal is the most common assumption given the presence of a small proportion of aberrant results. The <i>correct</i> data follow a normal distribution; aberrant results follow a different, usually much broader, distribution

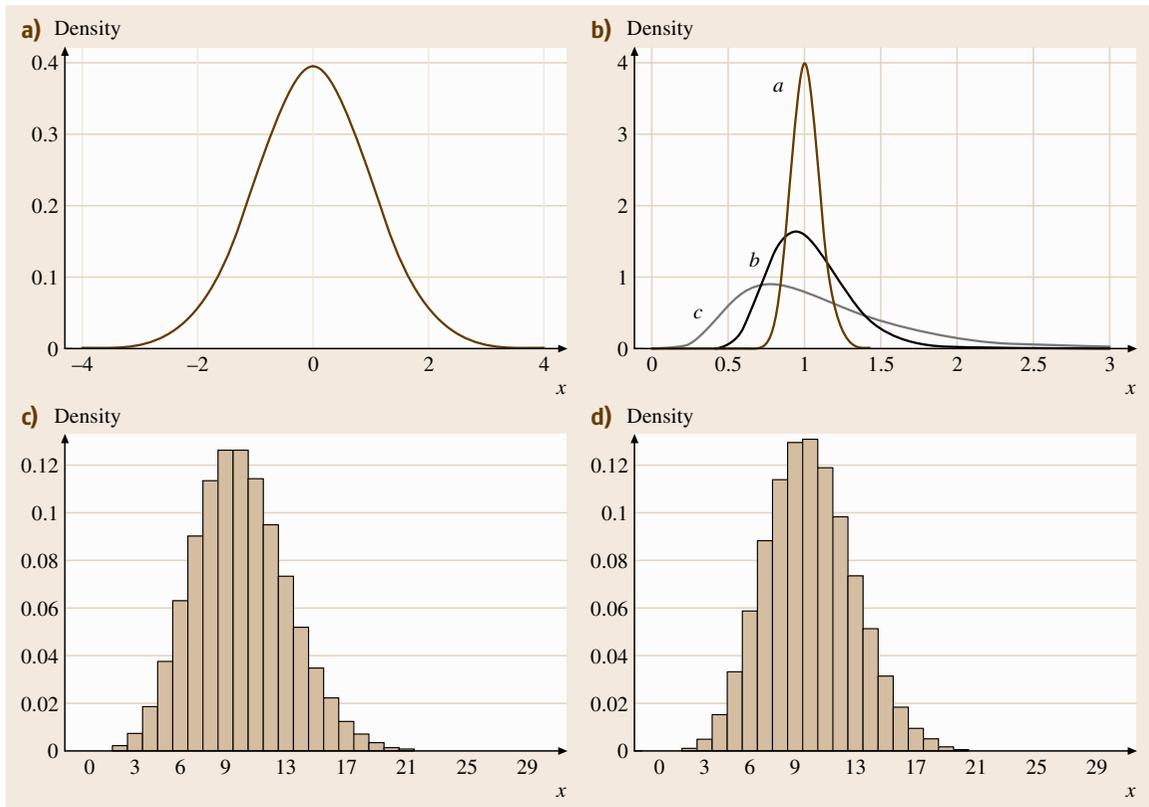


Fig. 3.4a-d Measurement data distributions. Figure 3.4 shows the *probability density function* for each distribution, not the probability; the area under each curve, or sum of discrete values, is equal to 1. Unlike probability, the probability density at a point x can be higher than 1. **(a)** The standard normal distribution (mean = 0, standard deviation = 1.0). **(b)** Lognormal distributions; mean on log scale = 0, standard deviation on log scale = a : 0.1, b : 0.25, c : 0.5. **(c)** Poisson distribution: lambda = 10. **(d)** Binomial distribution: 100 trials, $p(\text{success}) = 0.1$. Note that this provides the same mean as **(c)**

deviation any nonnegative value. The distribution is symmetric about the mean, and although the density falls off sharply, it is actually infinite in extent. The normal distribution arises naturally from the additive combination of many effects, even, according to the central limit theorem, when those effects do not themselves arise from a normal distribution. (This has an important consequence for means; errors in the mean of even three or four observations can often be taken to be normally distributed even where the parent distribution is not.) Furthermore, since small effects generally behave approximately additively, a very wide range of measurement systems show approximately normally distributed error.

- The lognormal distribution is closely related to the normal distribution; the logarithms of values from

a lognormal distribution are normally distributed. It most commonly arises when errors combine multiplicatively, instead of additively. The lognormal distribution itself is generally asymmetric, with positive skew. However, as shown in the figure, the shape depends on the ratio of standard deviation to mean, and approaches that of a normal distribution as the standard deviation becomes small compared with the mean. The simplest method of handling lognormally distributed data is to take logarithms and treat the logged data as arising from a normal distribution. As the standard deviation becomes small relative to the mean, the lognormal distribution tends towards the normal distribution.

- The Poisson and binomial distributions describe counts, and accordingly are *discrete* distributions;

they have nonzero density only for integer values of the variable. The Poisson distribution is applicable to cases such as radiation counting; the binomial distribution is most appropriate for systems dominated by sampling, such as the number of defective parts in a batch, the number of microbes in a fixed volume or the number of contaminated particles in a sample from an inhomogeneous mixture. In the limit of large counts, the binomial distribution tends to the normal distribution; for small probability, it tends to the Poisson distribution. Similarly, the Poisson distribution tends towards normality for small probability and large counts. Thus, the Poisson distribution is often a convenient approximation to the binomial, and as counts increase, the normal distribution can be used to approximate either.

Distributions Derived from the Normal Distribution. Before leaving the topic of distributions, it is important to be aware that other distributions are important in analyzing measurement data with normally distributed error. The most important for this discussion are

- the t -distribution, which describes the distribution of the means of small samples taken from a normal distribution. The t -distribution is routinely used for checking a method for significant bias or for comparing observations with limits,
- the chi-squared distribution, which describes inter alia the distribution of estimates of variance. Specifically, the variable $(n-1)s^2/\sigma^2$ has a chi-squared distribution with $\nu = n-1$ degrees of freedom. The chi-squared distribution is asymmetric with mean ν and variance 2ν ,
- the F -distribution, which describes the distribution of ratios of variances. This is important in comparing the spread of two different data sets, and is extensively used in analysis of variance as well as being useful for comparing the precision of alternative methods of measurement.

Probability and Significance

Given a particular distribution, it is possible to make predictions of the probability that observations will fall within a particular range. For example, in a normal distribution, the fraction of observations falling, by chance, within two standard deviations of the mean value is very close to 95%. This equates to the probability of an observation occurring in that interval. Similarly, the probability of an observation falling more than 1.65 standard deviations above the mean value is close to 5%.

These proportions can be calculated directly from the area under the curves shown in Fig. 3.4, and are available in tabular form, from statistical software and from most ordinary spreadsheet software.

Knowledge of the probability of a particular observation allows some statement about the significance of an observation. Observations with high probability of chance occurrence are not regarded as particularly significant; conversely, observations with a low probability of occurring by chance are taken as significant. Notice that an observation can only be allocated a probability if there is some assumption or hypothesis about the true state of affairs. For example, if it is asserted that the concentration of a contaminant is below some regulatory limit, it is meaningful to consider how likely a particular observation would be *given this hypothesis*. In the absence of any hypothesis, no observation is more likely than any other. This process of forming a hypothesis and then assessing the probability of a particular observation *given the hypothesis* is the basis of significance testing, and will be discussed in detail below.

3.3.2 Calculations and Software

Statistical treatment of data generally involves calculations, and often repetitive calculation. Frequently, too, best practise involves methods that are simply not practical manually, or require numerical solutions. Suitable software is therefore essential. Purpose-designed software for statistics and experimental design is widely available, including some free and open-source packages whose reliability challenges the best commercial software. Some such packages are listed in Sect. 3.12 at the end of this chapter. Many of the tests and graphical methods described in this short introduction are also routinely available in general-purpose spreadsheet packages. Given the wide availability of software and the practical difficulties of implementing accurate numerical software, calculations will not generally be described in detail. Readers should consult existing texts or software for further details if required.

However, it remains important that the software used is reliable. This is particularly true of some of the most popular business spreadsheet packages, which have proven notoriously inaccurate or unstable on even moderately ill-conditioned data sets. Any mathematical software used in a measurement laboratory should therefore be checked using typical measurement data to ensure that the numerical accuracy is sufficient. It may additionally be useful to test software using more

extreme test sets; some such sets are freely available (Sect. 3.12).

3.3.3 Statistical Methods

Graphical Methods

Graphical methods refer to the range of graphs or plots that are used to present and assess data visually. Some have already been presented; the histogram in Fig. 3.3 is an example. Graphical methods are easy to implement with a variety of software and allow a measurement scientist to identify anomalies, such as outlying data points or groups, departures from assumed distributions or models, and unexpected trends, quickly and with minimal calculation. A complete discussion of graphical methods is beyond the scope of this chapter, but some of the most useful, with typical applications, are presented below. Their use is strongly recommended in routine data analysis.

Figure 3.5 illustrates some basic plots appropriate for reviewing simple one-dimensional data sets. Dot

plots and strip charts are useful for reviewing small data sets. Both give a good indication of possible outliers and unusual clustering. Overinterpretation should be avoided; it is useful to gain experience by reviewing plots from random normal samples, which will quickly indicate the typical extent of apparent anomalies in small samples. Strip charts are simpler to generate (plot the data as the x variable with a constant value of y), but overlap can obscure clustering for even modest sets. The stacked dot plot, if available, is applicable to larger sets. Histograms become more appropriate as the number of data points increases. Box plots, or box-and-whisker plots (named for the lines extending from the rectangular box) are useful for summarizing the general shape and extent of data, and are particularly useful for grouped data. For example, the range of data from replicate measurements on several different test items can be reviewed very easily using a box plot. Box plots can represent several descriptive statistics, including, for example, a mean and confidence interval. However, they are most commonly based on quartiles. Traditionally,

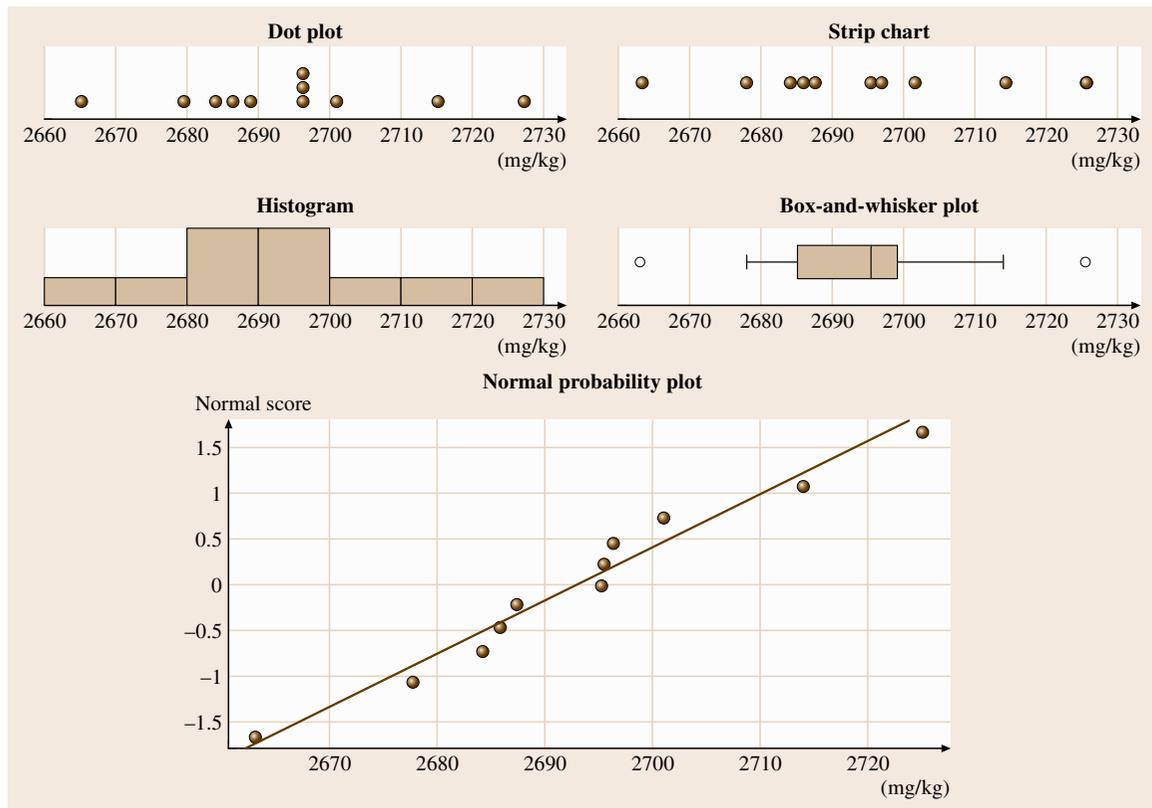


Fig. 3.5 Plots for simple data set review

the box extends from the first to the third quartile (that is, it contains the central 50% of the data points). The median is marked as a dividing line or other marker inside the box. The *whiskers* traditionally extend to the most distant data point within 1.5 times the interquartile range of the ends of the box. For a normal distribution, this would correspond to approximately the mean ± 2.7 standard deviations. Since this is just beyond the 99% confidence interval, more extreme points are likely to be outliers, and are therefore generally shown as individual points on the plot. Finally, a normal probability plot shows the distribution of the data plotted against the expected distribution assuming normality. In a normally distributed data set, points fall close to the diagonal line. Substantial deviations, particularly at either end of the plot, indicate nonnormality.

The most common graphical method for two-dimensional measurement data (such as measurand level/instrument response pairs) is a scatter plot, in which points are plotted on a two-dimensional space with dimensions corresponding to the dimensions of the data set. Scatter plots are most useful in reviewing data for linear regression, and the topic will accordingly be returned to below.

Planning of Experiments

Most measurements represent straightforward application of a measuring device or method to a test item. However, many experiments are intended to test for the presence or absence of some specific treatment effect – such as the effect of changing a measurement method or adjusting a manufacturing method. For example, one might wish to assess whether a reduction in preconditioning time had an effect on measurement results. In these cases, it is important that the experiment measures the intended effect, and not some external nuisance effect. For example, measurement systems often show significant changes from day to day or operator to operator. To continue the preconditioning example, if test items for short preconditioning were obtained by one operator and for long preconditioning by a different operator, operator effects might be misinterpreted as a significant conditioning effect. Ensuring that nuisance parameters do not interfere with the result of an experiment is one of the aims of good experimental design.

A second, but often equally important aim is to minimize the cost of an experiment. For example, a naïve experiment to investigate six possible effects might investigate each individually, using, say, three replicate measurements at each level for each effect: a total of

36 measurements. Careful experimental designs which vary all parameters simultaneously can, using the right statistical methods, reduce this to 16 or even 8 measurements and still achieve acceptable power.

Experimental design is a substantial topic, and a range of reference texts and software are available. Some of the basic principles of good design are, however, summarized below.

1. *Arrange experiments for cancellation*: the most precise and accurate measurements seek to cancel out sources of bias. For example, null-point methods, in which a reference and test item are compared directly by adjusting an instrument to give a zero reading, are very effective in removing bias due to residual current flow in an instrument. Simultaneous measurement of test item and calibrant reduces calibration differences; examples include the use of internal standards in chemical measurement, and the use of comparator instruments in gage block calibration. Difference and ratio experiments also tend to reduce the effects of bias; it is therefore often better to study differences or ratios of responses obtained under identical conditions than to compare absolute measurements.
2. *Control if you can; randomize if you cannot*: a good experimenter will identify the main sources of bias and control them. For example, if temperature is an issue, temperature should be controlled as far as possible. If direct control is impossible, the statistical analysis should include the nuisance parameter. *Blocking* – systematic allocation of test items to different strata – can also help reduce bias. For example, in a 2 day experiment, ensuring that every type of test item is measured an equal number of times on each day will allow statistical analysis to remove the between-day effect. Where an effect is known but cannot be controlled, and also to guard against unknown systematic effects, randomization should be used. For example, measurements should always be made in random order within blocks as far as possible (although the order should be recorded to allow trends to be identified), and test items should be assigned randomly to treatments.
3. *Plan for replication or obtaining independent uncertainty estimates*: without knowledge of the precision available, and more generally of the uncertainty, the experiment cannot be interpreted. Statistical tests all rely on comparison of an effect with some estimate of the uncertainty of the effect, usually based on observed precision. Thus, exper-

iments should always include some replication to allow precision to be estimated, or provide for additional information of the uncertainty.

4. *Design for statistical analysis: To consult a statistician after an experiment is finished is often merely to ask him to conduct a post-mortem examination. He can perhaps say what the experiment died of.* (R. A. Fisher, Presidential Address to the First Indian Statistical Congress, 1938). An experiment should always be planned with a specific method of statistical analysis in mind. Otherwise, despite the considerable range of tools available, there is too high a risk that no statistical analysis will be applicable. One particular issue in this context is that of *balance*. Many experiments test several parameters simultaneously. If more data are obtained on some combinations than others, it may be impossible to separate the different effects. This applies particularly to two-way or higher-order analysis of variance, in which interaction terms are not generally interpretable with unbalanced designs. Imbalance can be tolerated in some types of analysis, but not in all.

Significance Testing

General Principles. Because measurement results vary, there is always some doubt about whether an observed difference arises from chance variation or from an underlying, real difference. Significance testing allows the scientist to make reliable objective judgements on the basis of data gathered from experiments, while protecting against overinterpretation based on chance differences.

A significance test starts with some hypothesis about a true value or values, and then determines whether the observations – which may or may not appear to contradict the hypothesis – could reasonably arise by chance if the hypothesis were correct. Significance tests therefore involve the following general steps.

1. *State the question clearly, in terms of a null hypothesis and an alternate hypothesis:* in most significance testing, the null hypothesis is that there is no effect of interest. The alternate is always an alternative state of affairs such that the two hypotheses are mutually exclusive and that the combined probability of one or the other is equal to 1; that is, that no other situation is relevant. For example, a common null hypothesis about a difference between two values is: *there is no difference between the true val-*

ues ($\mu_1 = \mu_2$). The relevant alternate is that *there is a difference between the true values* ($\mu_1 \neq \mu_2$). The two are mutually exclusive (they cannot both be true simultaneously) and it is certain that one of them is true, so the combined probability is exactly 1.0. The importance of the hypotheses is that different initial hypotheses lead to different estimates of the probability of a contradictory observation. For example, if it is hypothesized that the (true) value of the measurand is exactly equal to some reference value, there is some probability (usually equal) of contradictory observations both above and below the reference value. If, on the other hand, it is hypothesized that the true value is less than or equal to the reference value, the situation changes. If the true value may be anywhere below or equal to the reference value, it is less likely that observations above the reference value will occur, because of the reduced chance of such observations from true values very far below the reference value. This change in probability of observations on one side or another must be reflected either in the choice of critical value, or in the method of calculation of the probability.

2. *Select an appropriate test:* different questions require different tests; so do different distribution assumptions. Table 3.2 provides a summary of the tests appropriate for a range of common situations. Each test dictates the method of calculating a value called the *test statistic* from the data.
3. *Calculate the test statistic:* in software, the test statistic is usually calculated automatically, based on the test chosen.
4. *Choose a significance level:* the significance level is the probability at which chance is deemed sufficiently unlikely to justify rejection of the null hypothesis. It is usually the measurement scientist's responsibility to choose the level of significance appropriately. For most common tests on measurement results, the significance level is set at 0.05,

Table 3.2 Common significance tests for normally distribution data. The following symbols are used: α is the desired significance level (usually 0.05); μ is the (true) value of the measurand; σ is the population standard deviation for the population described by μ (not that calculated from the data). \bar{a} is the observed mean; s is the standard deviation of the data used to calculate \bar{x} ; n is the number of data points. x_0 is the reference value; x_U, x_L are the upper and lower limits of a range. $\mu_1, \mu_2, \bar{x}_1, \bar{x}_2, s_1, s_2, n_1, n_2$ are the corresponding values for each of two sets of data to be compared ►

Test objective	Test name	Test statistic	Remarks
Tests on a single observed mean \bar{x} against a reference value or range			
Test for significant difference from the reference value x_0	Student <i>t</i> -test	$ x_0 - \bar{x} / (s / \sqrt{n})$	Hypothesis ($\mu = x_0$) against alternate ($x_0 \neq \mu$). Use a table of two-tailed critical values
Test for \bar{x} significantly exceeding an upper limit x_0	Student <i>t</i> -test	$(\bar{x} - x_0) / (s / \sqrt{n})$	Hypothesis ($\mu_1 = \mu_2$) against alternate ($\mu_1 > \mu_2$). Use a table of <i>one</i> -tailed critical values. Note that the sign of $x_0 - \bar{x}$ is retained
Test for \bar{x} falling significantly below a lower limit x_0	Student <i>t</i> -test	$(x_0 - \bar{x}) / (s / \sqrt{n})$	
Test for \bar{x} falling significantly outside a range [x_L, x_U]	Student <i>t</i> -test	$\max \left[\begin{array}{l} (x_L - \bar{x}) / (s / \sqrt{n}) \\ (\bar{x} - x_U) / (s / \sqrt{n}) \end{array} \right]$	Hypothesis: $x_L \leq \mu \leq x_U$ against the alternate $\mu < x_L, x_U < \mu$. Use a table of <i>one</i> -tailed critical values. This test assumes that the range is large compared with s , but $(x_U - x_L) > s$ gives adequate accuracy at the 5% significance level
Tests for significant difference between two means			
(a) With equal variance	Equal-variance <i>t</i> -test	$ \bar{x}_1 - \bar{x}_2 $	Hypothesis $\mu_1 = \mu_2$ against alternate ($\mu_1 \neq \mu_2$)
(b) With significantly different variance	Unequal-variance <i>t</i> -test		Use a table of two-tailed critical values. For equal variance, take degrees of freedom equal to $n_1 + n_2 - 2$. For unequal variance, take degrees of freedom equal to $\frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(n_1 - 1)/(s_1/n_1)^2 + (n_2 - 1)/(s_2/n_2)^2]}$
For testing the hypothesis $\mu_1 > \mu_2$ against the alternative $\mu_1 \leq \mu_2$, where μ_1 is the <i>expected</i> larger mean (not necessarily the larger observed mean), calculate the test statistic using $(\bar{x}_1 - \bar{x}_2)$ instead of $ \bar{x}_1 - \bar{x}_2 $ and use a one-tailed critical value			
Test n paired values for significant difference (constant variance)	Paired <i>t</i> -test	$\frac{ \bar{d} }{(s_d / \sqrt{n})}$, where $\bar{d} = \frac{1}{n} \sum_i x_{1,i} - x_{2,i}$ and $s_d = \frac{1}{n-1} \sum_i (x_{1,i} - x_{2,i})^2$	Hypothesis $\mu_d = 0$ against alternate $\mu_d \neq 0$. The sets must consist of pairs of measurements, such as measurements on the same test items by two different methods
Tests for standard deviations			
Test an observed standard deviation against a reference or required value σ_0	i) Chi-squared test	i) $(n-1)s^2/\sigma_0$	i) Compare $(n-1)s^2/\sigma_0$ with critical values for the chi-squared distribution with $n-1$ degrees of freedom
	ii) <i>F</i> -test	ii) s^2/σ_0	ii) Compare s^2/σ_0 with critical values for <i>F</i> for $(n-1)$ and infinite degrees of freedom For a test of $\sigma \leq \sigma_0$ against $\sigma > \sigma_0$, use the upper one-tailed critical value of chi-squared or <i>F</i> for probability α . To test $\sigma = \sigma_0$ against $\sigma \neq \sigma_0$, use two-tailed limits for chi-squared or compare $\max(s^2/\sigma_0, \sigma_0/s^2)$ against the upper one-tailed value for <i>F</i> for probability $\alpha/2$
Test for a significant difference between two observed standard deviations	<i>F</i> -test	s_{\max}^2/s_{\min}^2	Hypothesis: $\sigma_1 = \sigma_2$ against $\sigma_1 \neq \sigma_2$. s_{\max} is the larger observed standard deviation. Use the upper <i>one</i> -tailed critical value for <i>F</i> for a probability $\alpha/2$ using $n_1 - 1, n_2 - 1$ degrees of freedom
Test for one observed standard deviations s_1 significantly exceeding another (s_2)	<i>F</i> -test	s_1^2/s_2^2	Hypothesis: $\sigma_1 \leq \sigma_2$ against $\sigma_1 > \sigma_2$. Use the upper <i>one</i> -tailed critical value for <i>F</i> for a probability α using $n_1 - 1, n_2 - 1$ degrees of freedom
Test for homogeneity of variance among several groups of data	Levene's test	N/A	Levene's test is most simply estimated as a one-way analysis of variance performed on absolute values of group residuals, that is, $ x_{ij} - \hat{x}_j $, where \hat{x}_j is an estimate of the population mean of group j ; \hat{x}_j is usually the median, but the mean or another robust value can be used

or 5%. For stringent tests, 1% significance or less may be appropriate. The term *level of confidence* is an alternative expression of the same quantity; for example, the 5% level of significance is equal to the 95% level of confidence. Mathematically, the significance level is the probability of incorrectly rejecting the null hypothesis given a particular critical value for a test statistic (see below). Thus, one chooses the critical value to provide a suitable significance level.

5. *Calculate the degrees of freedom for the test:* the distribution of error often depends not only on the number of observations n , but on the number of degrees of freedom ν (Greek letter nu). ν is usually equal to the number of observations minus the number of parameters estimated from the data: $n - 1$ for a simple mean value, for example. For experiments involving many parameters or many distinct groups, the number of degrees of freedom may be very different from the number of observations. The number of degrees of freedom is usually calculated automatically in software.
6. *Obtain a critical value:* critical values are obtained from tables for the relevant distribution, or from software. Statistical software usually calculates the critical value automatically given the level of significance.
7. *Compare the test statistic with the critical value or examine the calculated probability (p -value).* Traditionally, the test is completed by comparing the calculated value of the test statistic with the critical value determined from tables or software. Usually (but not always) a calculated value higher than the critical value denotes significance at the chosen level of significance. In software, it is generally more convenient to examine the calculated probability of the observed test statistic, or p -value, which is usually part of the output. The p -value is always between 0 and 1; small values indicate a low probability of chance occurrence. Thus, if the p -value is below the chosen level of significance, the result of the test is significant and the null hypothesis is rejected.

Significance Tests for Specific Circumstances. Table 3.2 provides a summary of the most common significance tests used in measurement for normally distributed data. The calculations for the relevant test statistics are included, although most are calculated automatically by software.

Interpretation of Significance Test Results. While a significance test provides information on whether an observed difference could arise by chance, it is important to remember that statistical significance does not necessarily equate to practical importance. Given sufficient data, very small differences can be detected. It does not follow that such small differences are important. For example, given good precision, a measured mean 2% away from a reference value may be statistically significant. If the measurement requirement is to determine a value within 10%, however, the 2% bias has little practical importance.

The other chief limitation of significance testing is that a lack of statistical significance cannot prove the absence of an effect. It should be interpreted only as an indication that the experiment failed to provide sufficient evidence to conclude that there was an effect. At best, statistical insignificance shows only that the effect is not large compared with the experimental precision available. Where many experiments fail to find a significant effect, of course, it becomes increasingly safe to conclude that there is none.

Effect of Nonconstant Standard Deviation. Significance tests on means assume that the standard deviation is a good estimate of the population standard deviation and that it is constant with μ . This assumption breaks down, for example, if the standard deviation is approximately proportional to μ , a common observation in many fields of measurement (including analytical chemistry and radiological counting, although the latter would use intervals based on the Poisson distribution). In conducting a significance test in such circumstances, the test should be based on the best estimate of the standard deviation at the hypothesized value of μ , and not that at the value \bar{x} . To take a specific example, in calculating whether a measured value significantly exceeds a limit, the test should be based on the standard deviation at the limit, not at the observed value.

Fortunately, this is only a problem when the standard deviation depends very strongly on μ in the range of interest and where the standard deviation is large compared with the mean to be tested. For s/\bar{x} less than about 0.1, for example, it is rarely important.

Confidence Intervals

Statistical Basis of Confidence Intervals. A confidence interval is an interval within which a statistic (such as a mean or a single observation) would be expected to be observed with a specified probability.

Significance tests are closely related to the idea of confidence intervals. Consider a test for significant difference between an observed mean \bar{x} (taken from n values with standard deviation s) against a hypothesized measurand value μ . Using a t -test, the difference is considered significant at the level of confidence $1 - \alpha$ if

$$\frac{|\bar{x} - \mu|}{s/\sqrt{n}} > t_{\alpha, v, 2},$$

where $t_{\alpha, v, 2}$ is the two-tailed critical value of Student's t at a level of significance α . The condition for an *insignificant* difference is therefore

$$\frac{|\bar{x} - \mu|}{s/\sqrt{n}} \leq t_{\alpha, v, 2}.$$

Rearranging gives $|\bar{x} - \mu| \leq t_{\alpha, v, 2}s/\sqrt{n}$, or equivalently, $-t_{\alpha, v, 2}s/\sqrt{n} \leq \bar{x} - \mu \leq t_{\alpha, v, 2}s/\sqrt{n}$. Adding \bar{x} and adjusting signs and inequalities accordingly gives

$$\bar{x} - t_{\alpha, v, 2}s/\sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha, v, 2}s/\sqrt{n}.$$

This interval is called the $1 - \alpha$ *confidence interval* for μ . Any value of μ within this interval would be considered consistent with \bar{x} under a t -test at significance level α .

Strictly, this confidence interval cannot be interpreted in terms of the probability that μ is within the interval $\bar{x} \pm t_{\alpha, v, 2}s/\sqrt{n}$. It is, rather, that, in a long succession of similar experiments, a proportion $100(1 - \alpha)\%$ of the calculated confidence intervals would be expected to contain the true mean μ . However, because the significance level α is chosen to ensure that this proportion is reasonably high, a confidence interval does give an indication of the range of values that can reasonably be attributed to the measurand, based on the statistical information available so far. (It will be seen later that other information may alter the range of values we may attribute to the measurand.)

For most practical purposes, the confidence interval is quoted at the 95% level of confidence. The value of t for 95% confidence is approximately 2.0 for large degrees of freedom; it is accordingly common to use the range $\bar{x} \pm 2s/\sqrt{n}$ as an approximate 95% confidence interval for the value of the measurand.

Note that, while the confidence interval is in this instance symmetrical about the measured mean value, this is by no means always the case. Confidence intervals based on Poisson distributions are markedly asymmetric, as are those for variances. Asymmetric confidence intervals can also be expected when the standard deviation varies strongly with μ , as noted above in relation to significance tests.

Before leaving the topic of confidence intervals, it is worth noting that the use of confidence intervals is not limited to mean values. Essentially any estimated parameter estimate has a confidence interval. It is often simpler to compare some hypothesized value of the parameter with the confidence interval than to carry out a significance test. For example, a simple test for significance of an intercept in linear regression (below) is to see whether the confidence interval for the intercept includes zero. If it does, the intercept is not statistically significant.

Analysis of Variance

Introduction to ANOVA. Analysis of variance (ANOVA) is a general tool for analyzing data grouped by some factor or factors of interest, such as laboratory, operator or temperature. ANOVA allows decisions on which factors are contributing significantly to the overall dispersion of the data. It can also provide a direct measure of the dispersion due to each factor.

Factors can be qualitative or quantitative. For example, replicate data from different laboratories are grouped by the qualitative factor *laboratory*. This single-factor data would require one-way analysis of variance. In an experiment to examine time and temperature effects on a reaction, the data are grouped by both time and temperature. Two factors require two-way analysis of variance. Each factor treated by ANOVA must take two or more values, or *levels*. A combination of factor levels is termed a *cell*, since it forms a cell in a table of data grouped by factor levels. Table 3.3 shows an example of data grouped by time and temperature. There are two factors (time and temperature), and each has three levels (distinct values). Each cell (that is, each time/temperature combination) holds two observations.

The calculations for ANOVA are best done using software. Software can automate the traditional manual calculation, or can use more general methods. For example, simple grouped data with equal numbers of replicates within each cell are relatively simple to ana-

Table 3.3 Example data for two-way ANOVA

Time (min)	Temperature (K)		
	298	315	330
10	6.4	11.9	13.5
10	8.4	4.8	16.7
12	7.8	10.6	17.6
12	10.1	11.9	14.8
9	1.5	8.1	13.2
9	3.9	7.6	15.6

lyze using summation and sums of squares. Where there are different numbers of replicates per cell (referred to as an unbalanced design), ANOVA is better carried out by linear modeling software. Indeed, this is often the default method in current statistical software packages. Fortunately, the output is generally similar whatever the process used. This section accordingly discusses the interpretation of output from ANOVA software, rather than the process itself.

One-Way ANOVA. One-way ANOVA operates on the assumption that there are two sources of variance in the data: an effect that causes the true mean values of groups to differ, and another that causes data within each group to disperse. In terms of a statistical model, the i -th observation in the j -th group, x_{ij} , is given by

$$x_{ij} = \mu + \delta_j + \varepsilon_{ij},$$

where δ and ε are usually assumed to be normally distributed with mean 0 and standard deviations σ_b and σ_w , respectively. The subscripts “b” and “w” refer to the between-group effect and the within-group effect, respectively. A typical ANOVA table for one-way ANOVA is shown in Table 3.4 (The data analyzed are shown, to three figures only, in Table 3.5). The important features are

- The row labels, *Between groups* and *Within groups*, refer to the estimated contributions from each of the two effects in the model. The *Total* row refers to the total dispersion of the data.
- The columns “SS” and “df” are the sum of squares (actually, the sum of squared deviations from the relevant mean value) and the degrees of freedom for each effect. Notice that the total sum of squares and degrees of freedom are equal to the sum of those in the rows above; this is a general feature of ANOVA, and in fact the between-group SS and df can be calculated from the other two rows.
- The “MS” column refers to a quantity called the mean square for each effect. Calculated by dividing the sum of squares by the degrees of freedom, it can be shown that each mean square is an estimated

variance. The between-group mean square (MS_b) estimates $n_w\sigma_b^2 + \sigma_w^2$ (where n_w is the number of values in each group); the within-group mean square (MS_w) estimates the within-group variance σ_w^2 .

It follows that, if the between-group contribution were zero, the two mean squares should be the same, while if there were a real between-group effect, the between-group mean square would be larger than the within-group mean square. This allows a test for significance, specifically, a one-sided F -test. The table accordingly gives the calculated value for $F (= MS_b/MS_w)$, the relevant critical value for F using the degrees of freedom shown, and the p -value, that is, the probability that $F \geq F_{\text{calc}}$ given the null hypothesis. In this table, the p -value is approximately 0.08, so in this instance, it is concluded that the difference is not statistically significant. By implication, the instruments under study show no significant differences.

Finally, one-way ANOVA is often used for interlaboratory data to calculate repeatability and reproducibility for a method or process. Under interlaboratory conditions, repeatability standard deviation s_r is simply $\sqrt{MS_w}$.

The reproducibility standard deviation s_R is given by

$$s_R = \sqrt{\frac{MS_b + (n_w - 1)MS_w}{n_w}}.$$

Two-Way ANOVA. Two-way ANOVA is interpreted in a broadly similar manner. Each effect is allocated a row in an ANOVA table, and each main effect (that is, the effect of each factor) can be tested against the within-group term (often called the residual, or error, term in higher-order ANOVA tables). There is, however, one additional feature found in higher-order ANOVA tables: the presence of one or more *interaction terms*.

By way of example, Table 3.6 shows the two-way ANOVA table for the data in Table 3.3. Notice the *Interaction* row (in some software, this would be labeled *Time:Temperature* to denote which interaction it referred to). The presence of this row is best understood

Table 3.4 One-way ANOVA. Analysis of variance table

Source of variation	SS	df	MS	F	P-value	F _{crit}
Between groups	8.85	3	2.95	3.19	0.084	4.07
Within groups	7.41	8	0.93			
Total	16.26	11				

Table 3.5 One-way ANOVA. Data analyzed

Instrument			
A	B	C	D
58.58	59.89	60.76	61.80
60.15	61.02	60.78	60.60
59.65	61.40	62.90	62.50

by reference to a new statistical model

$$x_{ijk} = \mu + A_j + B_k + AB_{jk} + \varepsilon_{ijk} .$$

Assume for the moment that the factor A relates to the columns in Table 3.3, and the factor B to the rows. This model says that each level j of factor A shifts all results in column j by an amount A_j , and each level k of factor B shifts all values in row k by an amount B_j . This alone would mean that the effect of factor A is independent of the level of factor B . Indeed it is perfectly possible to analyze the data using the statistical model $x_{ijk} = \mu + A_j + B_k + \varepsilon_{ijk}$ to determine these main effects – even without replication; this is the basis of so-called two-way ANOVA without replication. However, it is possible that the effects of A and B are not independent; perhaps the effect of factor A depends on the level of B . In a chemical reaction, this is not unusual; the effect of time on reaction yield is generally dependent on the temperature, and vice versa. The term AB_{jk} in the above model allows for this, by associating a possible additional effect with every combination of factor levels A and B . This is the interaction term, and is the term referred to by the *Interaction* row in Table 3.6. If it is significant with respect to the within-group, or error, term, this indicates that the effects of the two main factors are not independent.

In general, in an analysis of data on measurement systems, it is safe to assume that the levels of the factors A and B are chosen from a larger possible population. This situation is analyzed, in two-way ANOVA, as a *random-effects model*. Interpretation of the ANOVA table in this situation proceeds as follows.

Table 3.6 Two-way ANOVA table

Source of variation	SS	df	MS	F	P-value	F _{crit}
Time	44.6	2	22.3	4.4	0.047	4.26
Temperature	246.5	2	123.2	24.1	0.0002	4.26
Interaction	15.4	4	3.8	0.8	0.58	3.63
Within	46.0	9	5.1			
Total	352.5	17	154.5			

1. Compare the interaction term with the within-group term.
2. If the interaction term is *not* significant, the main effects can be compared directly with the within-group term, as usually calculated in most ANOVA tables. In this situation, greater power can be obtained by pooling the within-group and interaction term, by adding the sums of squares and the degrees of freedom values, and calculating a new mean square from the new combined sum of squares and degrees of freedom. In Table 3.6, for example, the new mean square would be 4.7, and (more importantly) the degrees of freedom for the pooled effect would be 13, instead of 9. The resulting p -values for the main effects drop to 0.029 and 3×10^{-5} as a result. With statistical software, it is simpler to repeat the analysis omitting the interaction term, which gives the same results.
3. If the interaction term *is* significant, it should be concluded that, even if the main effects are not statistically significant in isolation, their combined effect is statistically significant. Furthermore, the effects are not independent of one another. For example, high temperature and long times might increase yield more than simply raising the temperature or extending the time in isolation. Second, compare the main effects with the interaction term (using an F -test on the mean squares) to establish whether each main effect has a statistically significant additional influence – that is, in addition to its effect in combination – on the results.

The analysis proceeds differently where both factors are *fixed effects*, that is, not drawn from a larger population. In such cases, all effects are compared directly with the within-group term.

Higher-order ANOVA models can be constructed using statistical software. It is perfectly possible to analyze simultaneously for any number of effects and all their interactions, given sufficient replication. However,

in two-way and higher-order ANOVA, some cautionary notes are important.

Assumptions in ANOVA. ANOVA (as presented above) assumes normality, and also assumes that the within-group variances arise from the same population. Departures from normality are not generally critical; most of the mean squares are related to sums of squares of group means, and as noted above, means tend to be normally distributed even where the parent distribution is nonnormal. However, severe outliers can have serious effects; a single severe outlier can inflate the within-group mean square drastically and thereby obscure significant main effects. Outliers can also lead to spurious significance – particularly for interaction terms – by moving individual group means. Careful inspection to detect outliers is accordingly important. Graphical methods, such as box plots, are ideal for this purpose, though other methods are commonly applied (see *Outlier detection* below).

The assumption of equal variance (homoscedasticity) is often more important in ANOVA than that of normality. Count data, for example, manifest a variance related to the mean count. This can cause seriously misleading interpretation. The general approach in such cases is to transform the data to give constant variance (not necessarily normality) for the transformed data. For example, Poisson-distributed count data, for which the variance is expected to be equal to the mean value, should be transformed by taking the square root of each value before analysis; this provides data that satisfies the assumption of homoscedasticity to a reasonable approximation.

Effect of Unbalanced Design. Two-way ANOVA usually assumes that the design is balanced, that is, all cells are populated and all contain equal numbers of observations. If this is not the case, the order that terms appear in the model becomes important, and changing the order can affect the apparent significance. Furthermore, the mean squares no longer estimate isolated effects, and comparisons no longer test useful hypotheses.

Advanced statistical software can address this issue to an extent, using various modified sums of squares (usually referred to as type II, III etc.). In practise, even these are not always sufficient. A more general approach is to proceed by constructing a linear model containing all the effects, then comparing the residual mean square with that for models constructed by omitting each main effect (or interaction term) in turn. Significant differences in the residual mean square indicate a significant effect, independently of the order of specification.

Least-Squares Linear Regression

Principles of Least-Squares Regression. Linear regression estimates the coefficients α_i of a model of the general form

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n,$$

where, most generally, each variable X is a *basis function*, that is, some function of a measured variable. Thus, the term covers both multiple regression, in which each X may be a different quantity, and polynomial regression, in which successive basis functions X are increasing powers of the independent variable (e.g., x , x^2 etc.). Other forms are, of course, possible. These all fall into the class of linear regression because they are linear in the coefficients α_i , not because they are linear in the variable X . However, the most common use of linear regression in measurement is to estimate the coefficients in the simple model

$$Y = \alpha_0 + \alpha_1 X,$$

and this simplest form – the form usually implied by the unqualified term *linear regression* – is the subject of this section.

The coefficients for the linear model above can be estimated using a surprisingly wide range of procedures, including robust procedures, which are resistant to the effects of outliers, and nonparametric methods, which make no distribution assumptions. In practise, by far the most common is simple least-squares linear regression, which provides the minimum-variance unbiased estimate of the coefficients when all errors are in the dependent variable Y and the error in Y is normally distributed. The statistical model for this situation is

$$y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i,$$

where ε_i is the usual error term and α_i are the true values of the coefficients, with estimated values a_i . The coefficients are estimated by finding the values that minimize the sum of squares

$$\sum_i w_i [y_i - (a_0 + a_1 x_i)]^2,$$

where the w_i are weights chosen appropriately for the variance associated with each point y_i . Most simple regression software sets the weights equal to 1, implicitly assuming equal variance for all y_i . Another common procedure (rarely available in spreadsheet implementations) is to set $w_i = 1/s_i^2$, where s_i is the standard deviation at y_i ; this *inverse variance weighting* is the correct weighting where the standard deviation varies significantly across the y_i .

The calculations are well documented elsewhere and, as usual, will be assumed to be carried out by software. The remainder of this section accordingly discusses the planning and interpretation of linear regression in measurement applications.

Planning for Linear Regression. Most applications of linear regression for measurement relate to the construction of a calibration curve (actually a straight line). The instrument response for a number of reference values is obtained, and the calculated coefficients a_i used to estimate the measurand value from signal responses on test items. There are two stages to this process. At the validation stage, the linearity of the response is checked. This generally requires sufficient power to detect departures from linearity and to investigate the dependence of precision on response. For routine measurement, it is sufficient to reestablish the calibration line for current circumstances; this generally requires sufficient uncertainty and some protection against erroneous readings or reference material preparation.

In the first, validation, study, a minimum of five levels, approximately equally spaced across the range of interest, are recommended. Replication is vital if a dependence of precision on response is likely; at least three replicates are usually required. Higher numbers of both levels and replication provide more power.

At the routine calibration stage, if the linearity is very well known over the range of interest and the intercept demonstrably insignificant, single-point calibration is feasible; two-point calibration may also be feasible if the intercept is nonzero. However, since there is then no possibility of checking either the internal consistency of the fit, or the quality of the fit, suitable quality control checks are essential in such cases. To provide additional checks, it is often useful to run a minimum of four to five levels; this allows checks for outlying values and for unsuspected nonlinearity. Of course, for extended calibration ranges, with less well-known linearity, it will be valuable to add further points. In the following discussion, it will be assumed that at least five levels are included.

Interpreting Regression Statistics. The first, and perhaps most important, check on the data is to inspect the fitted line visually, and wherever possible to check a residual plot. For unweighted regression (i. e., where $w_i = 1.0$) the residual plot is simply a scatter plot of the values $y_i - (a_0 + a_1x_i)$ against x_i . Where weighted regression is used, it is more useful to plot the weighted residuals $w_i[y_i - (a_0 + a_1x_i)]$. Figure 3.6 shows an ex-

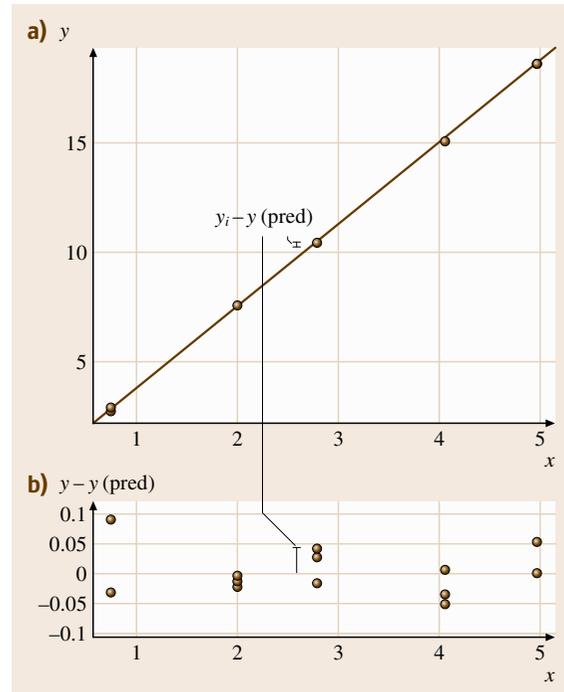


Fig. 3.6a,b Linear regression

ample, including the fitted line and data (Fig. 3.6a) and the residual plot (Fig. 3.6b). The residual plot clearly provides a much more detailed picture of the dispersion around the line. It should be inspected for evidence of curvature, outlying points, and unexpected changes in precision. In Fig. 3.6, for example, there is no evidence of curvature, though there might be a high outlying point at $x_i = 1$.

Regression statistics include the correlation coefficient r (or r^2) and a derived correlation coefficient r^2 (adjusted), plus the regression parameters a_i and (usually) their standard errors, confidence intervals, and a p -value for each based on a t -test for difference compared with the null hypothesis of zero for each.

The regression coefficient is always in the range -1 to 1 . Values nearer zero indicate a lack of linear relationship (not necessarily a lack of *any* relationship); values near 1 or -1 indicate a strong linear relationship. The regression coefficient will always be high when the data are clustered at the ends of the plot, which is why it is good practise to space points approximately evenly. Note that r and r^2 approach 1 as the number of degrees of freedom approaches zero, which can lead to overinterpretation. The adjusted r^2 value protects against this,

as it decreases as the number of degrees of freedom reduces.

The regression parameters and their standard errors should be examined. Usually, in calibration, the intercept a_0 is of interest; if it is insignificant (judged by a high p -value, or a confidence interval including zero) it may reasonably be omitted in routine calibration. The slope a_1 should always be highly significant in any practical calibration. If a p -value is given for the regression as a whole, this indicates, again, whether there is a significant linear relationship; this is usually well known in calibration, though it is important in exploratory analysis (for example, when investigating a possible effect on results).

Prediction from Linear Calibration. If the regression statistics and residual plot are satisfactory, the curve can be used for prediction. Usually, this involves estimating a value x_0 from an observation y_0 . This will, for many measurements, require some estimate of the uncertainty associated with prediction of a measurand value x from an observation y . Prediction uncertainties are, unfortunately, rarely available from regression software. The relevant expression is therefore given below.

$$s_{x_0} = \frac{s_{(y/x)}}{a_1} \left(w_0 + \frac{1}{n} + \frac{(y_0 - \bar{y}_w)^2}{a_1^2 \sum (w_i x_i^2 - n \bar{x}_w^2)} \right)^{1/2}$$

s_{x_0} is the *standard error of prediction* for a value x_0 predicted from an observation y_0 ; $s_{(y/x)}$ is the (weighted) residual standard deviation for the regression; \bar{y}_w and \bar{x}_w are the weighted means of the x and y data used in the calibration; n is the number of (x, y) pairs used; w_0 is a weighting factor for the observation y_0 ; if y_0 is a mean of n_0 observations, w_0 is $1/n_0$ if the calibration used unweighted regression, or is calculated as for the original data if weighting is used; s_{x_0} is the uncertainty arising from the calibration and precision of observation of y_0 in a predicted value x_0 .

Outlier Detection

Identifying Outliers. Measurement data frequently contain a proportion of extreme values arising from procedural errors or, sometimes, unusual test items. It is, however, often difficult to distinguish erroneous values from chance variations, which can also give rise to occasional extreme values. Outlier detection methods help to distinguish between chance occurrence as part of the normal population of data, and values that cannot reasonably arise from random variability.

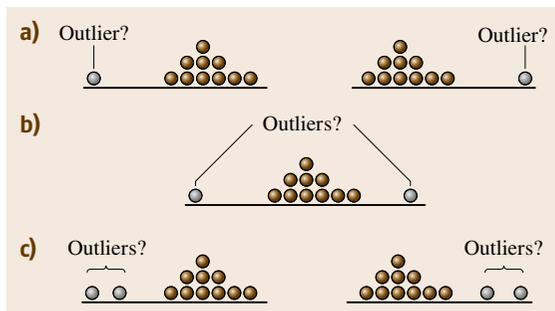


Fig. 3.7a-c Possible outliers in data sets

Graphical methods are effective in identifying possible outliers for follow-up. Dot plots make extreme values very obvious, though most sets have at least some apparent extreme values. Box-and-whisker plots provide an additional quantitative check; any single point beyond the whisker ends is unlikely to arise by chance in a small to medium data set. Graphical methods are usually adequate for the principal purpose of identifying data points which require closer inspection, to identify possible procedural errors. However, if critical decisions (including rejection – see below) are to be taken, or to protect against unwarranted follow-up work, graphical inspection should always be supported by statistical tests. A variety of tests are available; the most useful for measurement work are listed in Table 3.7. Grubb's tests are generally convenient (given the correct tables); they allow tests for single outliers in an otherwise normally distributed data set (Fig. 3.7a), and for simultaneous outlying pairs of extreme values (Fig. 3.7b, c), which would otherwise cause outlier tests to fail. Cochran's test is effective in identifying outlying variances, an important problem if data are to be subjected to analysis of variance or (sometimes) in quality control.

Successive application of outlier tests is permitted; it is not unusual to find that one exceptionally extreme value is accompanied by another, less extreme value. This simply involves testing the remainder of the data set after discovering an outlier.

Action on Detecting Outliers. A statistical outlier is only *unlikely* to arise by chance. In general, this is a signal to investigate and correct the cause of the problem. As a general rule, outliers should not be removed from the data set simply because of the result of a statistical test. However, many statistical procedures are seriously undermined by erroneous values, and long experience suggests that human error is the most common cause

Table 3.7 Tests for outliers in normally distributed data. The following assume an ordered set of data $x_1 \dots x_n$. Tables of critical values for the following can be found in ISO 5725:1995 part 2, among other sources. Symbols otherwise follow those in Table 3.2

Test objective	Test name	Test statistic	Remarks
Test for a single outlier in an otherwise normal distribution	i) Dixon's test	$n = 3 \dots 7 :$ $(x_n - x_{n-1}) / (x_n - x_1)$ $n = 8 \dots 10 :$ $(x_n - x_{n-1}) / (x_n - x_2)$ $n = 10 \dots 30 :$ $(x_n - x_{n-2}) / (x_n - x_3)$	The test statistics vary with the number of data points. Only the test statistic for a high outlier is shown; to calculate the test statistic for a low outlier, renumber the data in descending order. Critical values must be found from tables of Dixon's test statistic if not available in software
	ii) Grubb's test 1	$(x_n - \bar{x}) / s$ (high outlier) $(\bar{x} - x_1) / s$ (low outlier)	Grubb's test is simpler than Dixon's test if using software, although critical values must again be found from tables if not available in software
Test for two outliers on opposite sides of an otherwise normal distribution	Grubb's test 2	$1 - \left(\frac{(n-3)[s(x_3 \dots x_n)]^2}{(n-1)s^2} \right)$	$s(x_3 \dots x_n)$ is the standard deviation for the data excluding the two suspected outliers. The test can be performed on data in both ascending and descending order to detect paired outliers at each end. Critical values must use the appropriate tables
Test for two outliers on the same side of an otherwise normal distribution	Grubb's test 3	$(x_n - x_1) / s$	Use tables for Grubb's test 3
Test for a single high variance in l groups of data	Cochran's test	$C_{\bar{n}} = \frac{(s^2)_{\max}}{\sum_{i=1,l} s_i^2}$	$\bar{n} = \frac{1}{l} \sum_{i=1,l} n_i$

of extreme outliers. This experience has given rise to some general rules which are often used in processing, for example, interlaboratory data.

1. Test at the 95% and the 99% confidence level.
2. All outliers should be investigated and any errors corrected.
3. Outliers significant at the 99% level may be rejected unless there is a technical reason to retain them.
4. Outliers significant only at the 95% level should be rejected only if there is an additional, technical reason to do so.
5. Successive testing and rejection is permitted, but not to the extent of rejecting a large proportion of the data.

This procedure leads to results which are not unduly biased by rejection of chance extreme values, but are relatively insensitive to outliers at the frequency commonly encountered in measurement work. Note, however, that this objective can be attained without out-

lier testing by using robust statistics where appropriate; this is the subject of the next section.

Finally, it is important to remember that an outlier is only outlying in relation to some prior expectation. The tests in Table 3.7 assume underlying normality. If the data were Poisson distributed, for example, too many high values would be rejected as inconsistent with a normal distribution. It is generally unsafe to reject, or even test for, outliers unless the underlying distribution is known.

Robust Statistics

Introduction. Instead of rejecting outliers, robust statistics uses methods which are less strongly affected by extreme values. A simple example of a robust estimate of a population mean is the median, which is essentially unaffected by the exact value of extreme points. For example, the median of the data set (1, 2, 3, 4, 6) is identical to that of (1, 2, 3, 4, 60). The median, however, is substantially more variable than the mean when

the data are not outlier-contaminated. A variety of estimators have accordingly been developed that retain a useful degree of resistance to outliers without unduly affecting performance on normally distributed data. A short summary of the main estimators for means and standard deviations is given below. Robust methods also exist for analysis of variance, linear regression, and other modeling and estimation approaches.

Robust Estimators for Population Means. The median, as noted above, is a relatively robust estimator, widely available in software. It is very resistant to extreme values; up to half the data may go to infinity without affecting the median value. Another simple robust estimate is the so-called trimmed mean: the mean of the data set with two or more of the most extreme values removed. Both suffer from increases in variability for normally distributed data, the trimmed mean less so.

The mean suffers from outliers in part because it is a least-squares estimate, which effectively gives values a weight related to the square of their distance from the mean (that is, the *loss function* is quadratic). A general improvement can be obtained using methods which use a modified loss function. Huber (see Sect. 3.12 *Further Reading*) suggested a number of such estimators, which allocate a weight proportional to squared distance up to some multiple c of the estimated standard deviation \hat{s} for the set, and thereafter a weight proportional to distance. Such estimators are called M-estimators, as they follow from maximum-likelihood considerations. In Huber's proposal, the algorithm used is to replace each value x_i in a data set with z_i , where

$$z_i = \begin{cases} x_i & \text{if } \hat{X} - c \times \hat{s} < x_i < \hat{X} + c \times \hat{s} \\ \hat{X} \pm c \times \hat{s} & \text{otherwise} \end{cases},$$

and recalculate the mean \hat{X} , applying the process iteratively until the result converges. A suitable one-dimensional search algorithm may be faster. The estimated standard deviation is usually determined using a separate robust estimator, or (in Huber's proposal 2) iteratively, together with the mean. Another well-known approach is to use Tukey's biweight as the loss function; this also reduces the weight of extreme observations (to zero, for very extreme values).

Robust Estimators of Standard Deviation. Two common robust estimates of standard deviation are based on rank order statistics, such as the median. The first, the median absolute deviation (**MAD**), calculates the median of absolute deviations from the estimated mean

value \hat{x} , that is, median ($|x_i - \hat{x}|$). This value is not directly comparable to the standard deviation in the case of normally distributed data; to obtain an estimate of the standard deviation, a modification known as **MAD_e** should be used. This is calculated as **MAD**/0.6745. Another common estimate is based on the interquartile range (**IQR**) of a set of data; a normal distribution has standard deviation **IQR**/1.349. The **IQR** method is slightly more variable than the **MAD_e** method, but is usually easier to implement, as quartiles are frequently available in software. Huber's proposal 2 (above) generates a robust estimate of standard deviation as part of the procedure; this estimate is expected to be identical to the usual standard deviation for normally distributed data. **ISO 5725** provides an alternative iterative procedure for a robust standard deviation independently of the mean.

Using Robust Estimators. Robust estimators can be thought of as providing good estimates of the parameters for the good data in an outlier-contaminated set. They are appropriate when

- The data are expected to be normally distributed. Here, robust statistics give answers very close to ordinary statistics.
- The data are expected to be normally distributed, but contaminated with occasional spurious values, which are regarded as unrepresentative or erroneous. Here, robust estimators are less affected by occasional extreme values and their use is recommended. Examples include setting up quality control (**QC**) charts from real historical data with occasional errors, and interpreting interlaboratory study data with occasional problem observations.

Robust estimators are not recommended where

- The data are expected to follow nonnormal distributions, such as binomial, Poisson, chi-squared, etc. These generate extreme values with reasonable likelihood, and robust estimates based on assumptions of underlying normality are not appropriate.
- Statistics that represent the whole data distribution (including extreme values, outliers, and errors) are required.

3.3.4 Statistics for Quality Control

Principles

Quality control applies statistical concepts to monitor processes, including measurement processes, and

detect significant departures from normal operation. The general approach to statistical quality control for a measurement process is

1. regularly measure one or more typical test items (control materials),
2. establish the mean and standard deviation of the values obtained over time (ignoring any erroneous results),
3. use these parameters to set up warning and action criteria.

The criteria can include checks on stability of the mean value and, where measurements on the control material are replicated, on the precision of the process. It is also possible to seek evidence of emerging trends in the data, which might warn of impending or actual problems with the process.

The criteria can be in the form of, for example, permitted ranges for duplicate measurements, or a range within which the mean value for the control material must fall. Perhaps the most generally useful implementation, however, is in the form of a control chart. The following section therefore describes a simple control chart for monitoring measurement processes.

There is an extensive literature on statistical process control and control charting in particular, including a wide range of methods. Some useful references are included in Sect. 3.12 *Further Reading*.

Control Charting

A control chart is a graphical means of monitoring a measurement process, using observations plotted in a time-ordered sequence. Several varieties are in common use, including cusum charts (sensitive to sustained small bias) and range charts, which control precision. The type described here is based on a Shewhart mean chart. To construct the chart

- Obtain the mean \bar{x} and standard deviation s of at least 20 observations (averages if replication is used) on a control material. Robust estimates are recommended for this purpose, but at least ensure that no erroneous or aberrant results are included in this preliminary data.
- Draw a chart with date as the x -axis, and a y -axis covering the range approximately $\bar{x} \pm 4s$.
- Draw the mean as a horizontal line on the chart. Add two *warning limits* as horizontal lines at $\bar{x} \pm 2s$, and two further *action limits* at $\bar{x} \pm 3s$. These limits are approximate. Exact limits for specific probabil-

ities are provided in, for example, ISO 8258:1991 *Shewhart control charts*.

As further data points are accumulated, plot each new point on the chart. An example of such a chart is shown in Fig. 3.8.

Interpreting Control Chart Data

Two rules follow immediately from the action and warning limits marked on the chart.

- A point outside the action limits is very unlikely to arise by chance; the process should be regarded as out of control and the reason investigated and corrected.
- A point between the warning and action limits could happen occasionally by chance (about 4–5% of the time). Unless there is additional evidence of loss of control, no action follows. It may be prudent to remeasure the control material.

Other rules follow from unlikely sequences of observations. For example, two points outside the warning limits – whether on one side or alternate sides – is very unlikely and should be treated as actionable. A string of seven or more points above, or below, the mean – whether within the warning limits or not – is unlikely and may indicate developing bias (some recommendations consider ten such successive points as actionable). Sets of such rules are available in most textbooks on statistical process control.

Action on Control Chart *Action Conditions*

In general, actionable conditions indicate a need for corrective action. However, it is prudent to check that the control material measurement is valid before undertaking expensive investigations or halting a process. Taking a second control measurement is therefore advised, particularly for warning conditions. However, it is not sensible to continue taking control measurements until one falls back inside the limits. A single remeasurement is sufficient for confirmation of the out-of-control condition.

If the results of the second check do not confirm the first, it is sensible to ask how best to use the duplicate data in coming to a final decision. For example, should one act on the second observation? Or perhaps take the mean of the two results? Strictly, the correct answer requires consideration of the precision of the means of duplicate measurements taken over the appropriate time interval. If this is available, the appropriate limits can

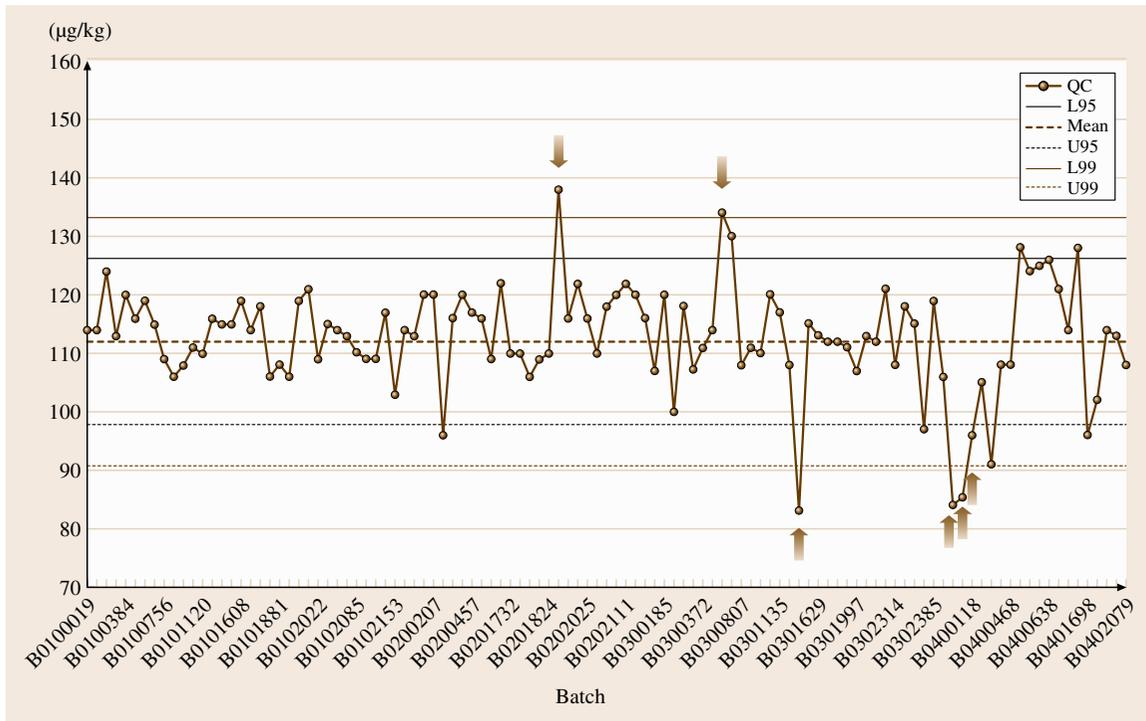


Fig. 3.8 QC chart example. The figure shows successive QC measurements on a reference material certified for lead content. There is evidence of loss of control at points marked by arrows

be calculated from the relevant standard deviation. If not, the following procedure is suggested: First, check whether the difference between the two observations is consistent with the usual operating precision (the results should be within approximately $2.8s$ of one another). If so, take the mean of the two, and compare this with new

limits calculated as $\bar{x} \pm 2s/\sqrt{2}$ and $\bar{x} \pm 3s/\sqrt{2}$ (this is conservative, in that it assumes complete independence of successive QC measurements; it errs on the side of action). If the two results do not agree within the expected precision, the cause requires investigation and correction in any case.

3.4 Uncertainty and Accuracy of Measurement and Testing

3.4.1 General Principles

In metrology and testing, the result of a measurement should always be expressed as the measured quantity value together with its uncertainty. The uncertainty of measurement is defined as a nonnegative parameter characterizing the dispersion of the quantity values being attributed to a measurand [3.17].

Measurement accuracy, which is the closeness of agreement between a measured quantity value and the true quantity value of a measurand, is a positive formulation for the fact that the measured value

is deviating from the true value, which is considered unique and, in practise, unknowable. The deviation between the measured value and the true value or a reference value is called the measurement error.

Since the 1990s there has been a conceptual change from the traditionally applied *error approach* to the *uncertainty approach*.

In the *error approach* it is the aim of a measurement to determine an *estimate of the true value* that is as close as possible to that single true value. In the *uncertainty approach* it is assumed that the information from

measurement only permits assignment of an *interval of reasonable values* to the measurand.

The master document, which is acknowledged to apply to all measurement and testing fields and to all types of uncertainties of quantitative results, is the *Guide to the Expression of Uncertainty in Measurement (GUM)* [3.19]. The Joint Committee for Guides in Metrology Working Group 1 (JCGM-WG1), author of the *GUM*, is producing a complementary series of documents to accompany the *GUM*.

The *GUM uncertainty philosophy* has already been introduced in Chap. 1, its essential points are

- A measurement quantity X , of which the true value is not known exactly, is considered as a stochastic variable with a probability function. Often it is assumed that this is a normal (*Gaussian*) distribution.
- The result x of a measurement is an estimate of the expectation value $E(X)$ for X .
- The standard uncertainty $u(x)$ of this measured value is equal to the square root of the variance $V(X)$.
- Expectation (quantity value) and variance (standard uncertainty) are estimated either
 - by statistical processing of repeated measurements (*type A uncertainty evaluation*) or
 - by other methods (*type B uncertainty evaluation*).
- The result of a measurement has to be expressed as a quantity value together with its uncertainty, including the unit of the measurand.

The methodology of measurement evaluation and determination of measurement uncertainty are compiled

in Fig. 3.9. The statistical evaluation of results has been described in detail in Sect. 3.3.

3.4.2 Practical Example: Accuracy Classes of Measuring Instruments

All measurements of quantity values for single measurands as well as for multiple measurands need to be performed with appropriate measuring instruments, devices for making measurements, alone or in conjunction with one or more supplementary devices.

The quality of measuring instruments is often defined through *limits of errors* as description of the *accuracy*.

Accuracy classes are defined [3.17] as classes of measuring instruments or measuring systems that meet stated metrological requirements that are intended to keep measurement errors or instrumental measurement uncertainties within specified limits under specified operating conditions. An accuracy class is usually denoted by a number or symbol adopted by convention. Analog measuring instruments are divided conventionally into accuracy classes of 0.05, 0.1, 0.2, 0.3, 0.5, 1, 1.5, 2, 2.5, 3, and 5. The accuracy classes p represent the maximum permissible relative measurement error in %. For example an accuracy class of 1.0 indicates that the limits of error – in both directions – should not exceed 1% of the full-scale deflection. In digital instruments, the limit of indication error is ± 1 of the least significant unit of the digital indication display.

In measuring instruments with an analog indication, the measured quantity is determined by the position

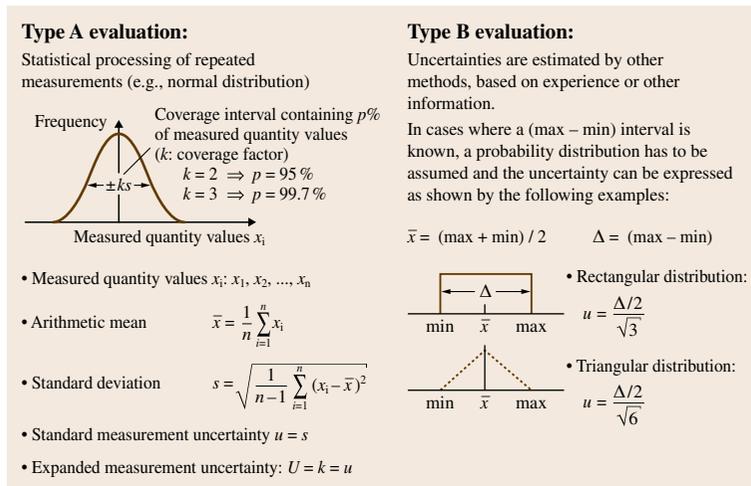
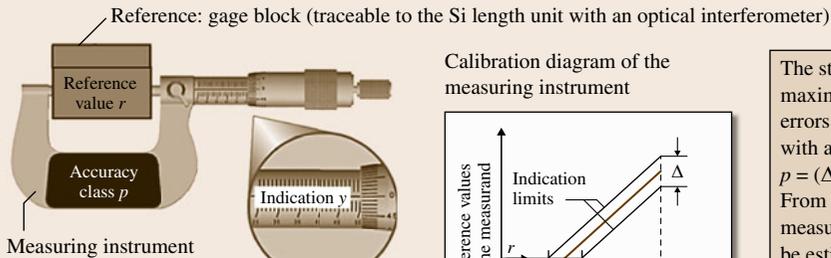


Fig. 3.9 Principles of measurement evaluation and determination of uncertainty of measurement for a single measurand x

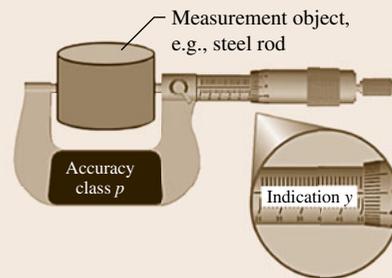
Measurement uncertainty of a single measurand with a single measuring instrument

Example from dimensional metrology

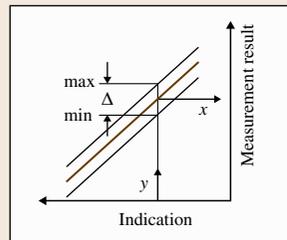
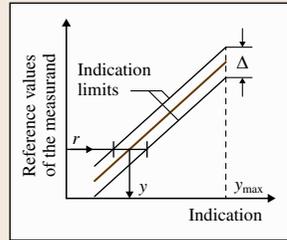
(1) Calibration of measuring instrument (measurand: length)



(2) Measurement



Calibration diagram of the measuring instrument



The strip Δ is the range of the maximum permissible measurement errors of a measuring instrument with an accuracy class $p = (\Delta / (2y_{\max})) \cdot 100$ [%]. From Δ or p , the instrument measurement uncertainty u_{instr} can be estimated in a type B evaluation. Assuming a rectangular distribution (Fig. 3.9) it follows that $u_{\text{instr}} = (\Delta/2)\sqrt{3}$, or $u_{\text{instr}} = ((p/100) \cdot y_{\max}) / \sqrt{3}$. The relative instrument measurement uncertainty [%] $\delta_{\text{instr}} = u_{\text{instr}} / u_{\text{max}}$ is given by $\delta_{\text{instr}} = p / \sqrt{3}$.

Measurement result:
Quantity value x
 \pm instrument measurement uncertainty u_{instr} .

Fig. 3.10 Method for obtaining a measurement result and estimating the instrument measurement uncertainty

of the indicator on the scale. The limits of errors (percentages) are usually given at the full-scale amplitude (maximum value of measurement range). From the accuracy class p , also the instrumental measurement uncertainty u_{instr} can be estimated. In Fig. 3.10, the method for obtaining a measurement result and measurement uncertainty for a single measurand with a single measuring instrument is shown.

As illustrated in Fig. 3.10, a measuring instrument gives as output an *indication*, which has to be related to the quantity value of the measurand through a calibration diagram. A calibration diagram represents the relation between indications of a measuring instrument and a set of reference values of the measurand. At the maximum indication value (maximum measurement range) y_{\max} the width Δ of the strip of the calibration diagram is the range of the maximum permissible measurement errors.

From the determination of Δ the accuracy class p in % follows as

$$p = \left(\frac{\Delta}{2y_{\max}} \right) \cdot 100 \text{ [%]}.$$

Note that, at indicator amplitudes lower than the maximum y_{\max} , the actual relative maximum permissible measurement errors p_{act} for the position y_{act} on the scale need to be determined as

$$p_{\text{act}} = p \cdot \left(\frac{y_{\max}}{y_{\text{act}}} \right).$$

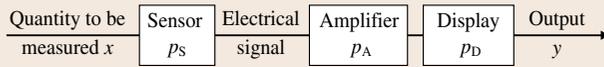
For the estimation of the standard measurement uncertainty it can be considered in an uncertainty estimation of type B that all values in the range between the limits of indications have the same probability – as long as no other information is available. This kind of distribution is called a rectangular distribution (Fig. 3.9). Therefore, the standard uncertainty is equal to

$$u_{\text{instr}} = \frac{(\Delta/2)}{\sqrt{3}} = \frac{((p/100) \cdot y_{\max})}{\sqrt{3}}.$$

Example 3.1: What is the measurement uncertainty of a measurement result obtained by a measurement with an analog voltmeter (accuracy class 2.0) with a maximum amplitude of 380 V, when the indicator is at 220 V?

Measurement uncertainty of a measurement system or measurement chain

Consider a measurement system, consisting in the simplest case of three components, namely a sensor (accuracy class p_S), an amplifier (accuracy class p_A) and a display (accuracy class p_D)



The measurement uncertainty of the system can be estimated by applying the law of the propagation of uncertainties (see Sect. 3.4.3)

$$u_{\text{System}}/|y| = \sqrt{(u_S^2/x_S^2 + u_A^2/x_A^2 + u_D^2/x_D^2)},$$

where $u_S/x_S + u_A/x_A, u_D/x_D$, are the relative instrument uncertainties of sensor, amplifier and display, which can be expressed by their accuracy classes as $p_S/\sqrt{3}, p_A/\sqrt{3}, p_D/\sqrt{3}$.

It follows that
$$u_{\text{System}}/|y| = \frac{\sqrt{(p_S^2 + p_A^2 + p_D^2)}}{\sqrt{3}}$$

For a measurement system of n components in line, the following formula characterizes the relative *uncertainty budget* of the measuring chain

$$\delta_{\text{chain}} = u_{\text{chain}}/|y| = \frac{\sqrt{(\sum p_i^2)}}{\sqrt{3}} \quad (i = 1 \dots n)$$

Consideration: actual relative maximum permissible measurement errors for 220 V and limits of error expressed in measurement units (V as scale divisions) are

$$p_{220,\text{rel}} = 2.0\% \cdot \frac{380 \text{ V}}{220 \text{ V}} = 3.5\% ;$$

$$p_{\text{abs}} = 380 \text{ V} \cdot \frac{2.0\%}{100\%} = 7.6 \text{ V} \quad (\text{limits of error})$$

$$u_{\text{instr,rel}} = \frac{p_{\text{rel}}}{\sqrt{3}} = \frac{3.5\%}{\sqrt{3}} = 2.0\% \quad \text{and}$$

$$u_{\text{instr,abs}} = \frac{p_{\text{abs}}}{\sqrt{3}} = \frac{7.6 \text{ V}}{\sqrt{3}} = 4.4 \text{ V} .$$

It is obvious that the relative standard uncertainties are smallest at y_{max} .

Since a rectangular distribution was assumed, it is not reasonable to apply the coverage factor k , because this approach assumes a Gaussian distribution. Instead, the standard uncertainty u_{instr} should be stated. It normally suffices to report the uncertainties to at most two significant digits – and also to provide information on how it was determined. Finally, the measurement uncertainty allows the experimenter to decide whether the used instrument is appropriate for his/the customer's needs.

Answer: The result of the example could be reported as 220 V \pm 4.4 V. The measurement uncertainty of the

Fig. 3.11 Method for estimating the measurement system uncertainty

instrument is reported as a standard uncertainty (coverage factor $k = 1$) and was obtained by type B evaluation only considering the instrument accuracy class.

If instead of a single measuring instrument, a measuring system or a measuring chain is used, consisting in the simplest case of a sensor, an amplifier, and a display, the accuracy classes of the components of the measuring system can also be used to estimate the instrumental system uncertainty, as illustrated in Fig. 3.11.

3.4.3 Multiple Measurement Uncertainty Components

The method outlined in Figs. 3.9 and 3.10 considers only one single measurement quantity and only the sources covered by only one variable. However, very often uncertainty evaluations have to be related to functional combinations of measured quantities or uncertainty components $y = f(x_1, x_2, x_3, \dots, x_n)$. In these cases, for uncorrelated (i. e., independent) values, the single uncertainties are combined by applying the law of propagation of uncertainty to give the so-called combined measurement uncertainty

$$u_{\text{combined}}(y) = \sqrt{\sum \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i)} .$$