

Springer Series in Solid and Structural Mechanics 4

Alessandro Freddi
Giorgio Olmi
Luca Cristofolini

Experimental Stress Analysis for Materials and Structures

Stress Analysis Models for Developing
Design Methodologies

 Springer

Springer Series in Solid and Structural Mechanics

Volume 4

Series editors

Michel Frémond, Rome, Italy

Franco Maceri, Rome, Italy

More information about this series at <http://www.springer.com/series/10616>

Alessandro Freddi · Giorgio Olmi
Luca Cristofolini

Experimental Stress Analysis for Materials and Structures

Stress Analysis Models for Developing
Design Methodologies

 Springer

Alessandro Freddi
Department of Industrial Engineering
University of Bologna
Bologna
Italy

Luca Cristofolini
Department of Industrial Engineering
University of Bologna
Bologna
Italy

Giorgio Olmi
Department of Industrial Engineering
University of Bologna
Bologna
Italy

ISSN 2195-3511 ISSN 2195-352X (electronic)
Springer Series in Solid and Structural Mechanics
ISBN 978-3-319-06085-9 ISBN 978-3-319-06086-6 (eBook)
DOI 10.1007/978-3-319-06086-6

Library of Congress Control Number: 2015932851

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2015

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

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

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

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media
(www.springer.com)

Nichts ist so praktisch wie eine gute Theorie

—Kurt Lewin (1890–1947)

(There is nothing so practical as a good theory)

but:

Nothing helps scientific thought like a good Experiment.

Engineering and Physics cannot be thought of without models which represent the real world to the best of our knowledge.

—Friedrich Pfeiffer, Hartmut Bremer
(2015) CISM–AIMETA–Udine.

Since the birth of modern science, the hypothesis as well as its experimental verification have a mathematical nature, both being expressed by a set of physical measurable—not qualitative—quantities.....

—Emanuele Severino (1984)
Modern Philosophy
Rizzoli Inc., Italy

Foreword

A model is an abstract mathematical description of a more or less complex entity, which is important in engineering sciences to investigate and predict the entity behavior in relation to performance, reliability, and safety. Correct modeling of a technical entity, whether material, structure, device, equipment or system, is thus fundamental to engineering purposes. This holds, in particular, for mechanical structures or systems, for which experiments to support modeling are often unavoidable, following the definition of inverse problem from the measurement of effects to identification of causes given in the preface of this book. In these cases, design of experiment is an essential tool to verify, develop, or refine models. Experiment and experience greatly help in finding simplifications making the model just as complex as necessary to contain all relevant parameters. However, in model building, physical and statistical validation of model assumptions should precede data analysis. In this book Professor Freddi, supported by two of his closest collaborators, has condensed the experiences gained in a long academic activity in this field in cooperation with the industry. The book gives a careful introduction to methods and tools used in Experimental Stress Analysis, opening the door to new investigations on failure mechanisms it is useful for mechanical engineers in practice as well as a textbook, not least for the large number of solved practical examples.

Firenze and Zürich, April 2015

Alessandro Birolini
Ph.D. Professor Emeritus of Reliability
Engineering at the ETH Zurich

Preface

The contents presented in this book are, in the opinion of the authors, an essential conceptual basis for the training of professional mechanical engineers in the field of experimental methods applied to theory of structures. Our intention is not to give an overview of all the experimental methods used in stress analysis, as we present only those that we have utilized for research purposes and scientific and technical consulting. Experimental analysis methods are based on different branches of science such as mechanics, electronics, optics, information theory, etc.: excellent specialized books are available written by specialists from different areas, which focus exclusively on methods. This book is a complementary reading to such specialized works, in that it gives an outline of some methods and shows how their applications make it possible to understand the *foundation* of a problem even before obtaining a detailed solution. For this latter task, the computing methods that cover nearly every need in solving particular cases are the best.

While historically experimental analysis was developed as a surreptitious tool due to a lack of analytical solutions to structural problems, today it is an instrument for clarifying the limitations of analytical theories, but primarily, due to the inverse nature of experimentation, for identifying unknown parameters, *integrating experimental data into analytical models*. Experimental Stress Analysis was classically regarded as a collection of experimental methods dedicated to the measurement of deformations of loaded bodies, and then to the corresponding states of stress. This has been modified into a methodology to build acceptable models for setting up phenomenological theories and supporting design practices.

It is common to speak of design for safety, reliability, or design to prevent high cycle fatigue, low cycle fatigue, crack propagation, multi-mode failures, etc. On the basis of years of teaching and consulting experience, we believe that the ultimate goal of experimental analysis is not only the knowledge of a state of stress but also the design and assessment of the integrity of structural systems.

There is another issue here that must be dealt with: Experimental Stress Analysis requires a variety of devices, testing machines, etc. that may be available on the market, often packaged in the form of black boxes. Self-made equipments give

many more opportunities for adjusting to specific problems and have an unparalleled role for training the experimentalist and also for education in structural design. The laboratory practice is highly educational for this purpose since it is conceptually similar to the practice of design. Both deal with problems in which a lot of data is unknown. This book is therefore oriented to applications, which require a self-made laboratory equipment on the assumption that it may offer a useful aid to help graduate students develop sensitivity to quickly discovering the few controlling variables and the essential tests for solving the problem.

According to its inverse methodology nature, experimental analysis has recently found new tasks in the solution of reliability problems. The estimation of reliability, and therefore also the answer to the question of a possible extension of life, requires experiments in order to identify the causes of defects and failures and to measure the failure rate. Therefore, the role of experimentation is not one of simple verification of a theoretical prediction, but is a continuous monitoring and control of processes and an invaluable tool for the estimation of the life of technical systems.

Finally, we did not use the more accredited and common label of Experimental Mechanics as the title to cover the topics of this book. Generic nominalization might lead to a confused definition of boundaries and to misunderstandings: important topics, such as Experimental Vibrations Analysis, have for many years constituted disciplines in their own right and are not covered in this book. Along the same line, there is no reason to limit Experimental Mechanics to solid bodies.

References to experimental stress analysis can be found in materials produced by scientific societies in various countries, such as SEM (American Association of Experimental Mechanics) in the USA, GESA (Gesellschaft für Experimentelle Spannung Analyse) in Germany, GAMAC (Avancement des Methodes d'Analyse des Contraintes) in France, BSSM (British Society for Strain Measurement) in the UK, AIAS (Italian Association of Stress Analysis) in Italy, EURASEM (European Association for Experimental Mechanics) for Europe, the Danubia-Adria Symposium for Central Europe, and the International Committee IMEKO TC 15 (Experimental Mechanics) with related YSESM, (Youth Symposium for Experimental Solid Mechanics) expressly devoted to young researchers.

Bologna, April 2015

Alessandro Freddi
Giorgio Olmi
Luca Cristofolini

Contents

Part I A Brief Review of the Experimental Methods Utilized in the Book

1	Introduction to Inverse Problems	3
1.1	Premise	3
1.1.1	General Rules	4
1.1.2	Rules for Inverse Problems	4
1.2	Forward and Inverse Problems for Elastic Discretized Structures	5
1.2.1	Quality Indicators of Inverse Solutions	13
1.2.2	Inverse Solution for Systems in Matrix Form	14
1.3	Systems in Functional Form	17
1.3.1	Regularization Method of Tikhonov-Phillips	18
1.3.2	Regularization Using Regularization Matrix	19
1.3.3	Further Reading	20
	References	20
2	Introduction to the Application of Strain Gages	23
2.1	Properties of Strain Gages	23
2.1.1	Relationship Between Strain and Resistance Change	24
2.1.2	Materials for Metal Strain Gages	26
2.1.3	Resistance Values	28
2.1.4	Transverse Sensitivity and Strain Gage Factor	32
2.1.5	Influence of a Temperature Variation	34
2.1.6	Compensation for Thermal Output	36
2.2	Strain Gage Rosettes	38
2.2.1	Three-Gage Rectangular ($0^\circ 45^\circ 90^\circ$) Rosettes	39
2.2.2	Three-Gage ($0^\circ 120^\circ 240^\circ$) or ($0^\circ 60^\circ 120^\circ$) Rosettes	41

- 2.3 Potentiometric Circuit 43
 - 2.3.1 Measurement of Dynamic Strains 44
- 2.4 Wheatstone Bridge 48
 - 2.4.1 Shunt Calibration 50
 - 2.4.2 Bridge Excitation 52
- 2.5 Bridge Configurations 53
 - 2.5.1 Quarter Bridge 53
 - 2.5.2 Half Bridge 54
 - 2.5.3 Full Bridge 55
 - 2.5.4 Eliminating Cable Effects with Three-Wire Circuit . . . 57
- 2.6 Semiconductor Strain Gage 58
- 2.7 Tests on Piezo-Resistive Hydrogenated Amorphous Silicon Strain Gages 58
 - 2.7.1 Gage Preparation and Testing 59
 - 2.7.2 Calibration and Test Results 61
 - 2.7.3 Temperature Response 63
- 2.8 Load Cells with Strain Gages Sensors 64
 - 2.8.1 Bending Load Cells 65
 - 2.8.2 Ring Load Cells 65
 - 2.8.3 Shear Load Cells 67
- 2.9 Load Cell Calibration 69
- 2.10 Load Cells with More Degrees of Freedom 70
 - 2.10.1 Load Cell for Three Forces 73
 - 2.10.2 Load Cell for Two Forces and One Torsional Moment 80
 - 2.10.3 Load Cells at Six Degrees of Freedom 87
 - 2.10.4 Regularization of the Six-Degrees Load Cell 92
 - 2.10.5 Determination of the Damping Factor 95
- References 97

- 3 Introduction to Photoelasticity 101**
 - 3.1 Premise 101
 - 3.2 Nature of Light 101
 - 3.2.1 Polarized Light 104
 - 3.2.2 Plane Polarized Light 104
 - 3.2.3 Circularly Polarized Light 106
 - 3.2.4 Elliptically Polarized Light 107
 - 3.2.5 Plane Polarizer 108
 - 3.2.6 Double Refraction or Birefringence 109
 - 3.3 Light Treatment in an Optical System 111
 - 3.3.1 Optical System with a Polarizer and a Birefringent Filter 112
 - 3.4 Light Intensity 114
 - 3.5 Optics of Photoelasticity 114

- 3.6 Polariscope 116
- 3.7 Plane Polariscopescope 118
 - 3.7.1 Polarizer and Analyzer with Crossed Axes 118
- 3.8 Circular Polariscopescope 120
 - 3.8.1 Polarizer Along y Axis and Analyzer
Along x Axis 120
 - 3.8.2 Both Polarizer and Analyzer Along the y Axis 122
- 3.9 Isochromatics 122
- 3.10 Isoclines 126
 - 3.10.1 Isotropic and Singular Points 127
 - 3.10.2 Determination of Fractional Fringe Orders 128
 - 3.10.3 Measurement of Fractional Orders by Means
of Quarter-Wave Plates 132
 - 3.10.4 Drawing Stress Trajectories from Isoclines 136
- 3.11 Principal Stresses Separation 139
 - 3.11.1 Shear Stresses and Normal Stresses Difference
Determination 139
 - 3.11.2 Determination of Stress Sign 139
 - 3.11.3 Integration of Equilibrium Equations 141
 - 3.11.4 Overview of Experimental Methods for Principal
Stresses Separation 142
- 3.12 Materials for Photoelastic Models 144
 - 3.12.1 Boundary Effects 145
 - 3.12.2 Stress Freezing 146
 - 3.12.3 Calibration of Photoelastic Material 146
 - 3.12.4 Material Sensitivity 149
- 3.13 Three-Dimensional Photoelasticity: The Frozen
Stress Method 150
 - 3.13.1 Limit of Three-Dimensional Photoelasticity 151
 - 3.13.2 Overview of the Mold-Making
and Casting Technique 152
 - 3.13.3 Stress State in Plane External Surfaces
and in Planes Perpendicular to the External
Surface 154
 - 3.13.4 Determination of Stress State at Points
Inside the Model 155
- 3.14 Birefringent Coating Method 159
 - 3.14.1 Measurement of Principal Strains Directions 159
 - 3.14.2 Strains and Stresses Measurement 160
 - 3.14.3 Coating Calibration 161
- References 164

4	Introduction to Holographic Interferometry	167
4.1	Holography	167
4.1.1	The Laser Light	168
4.1.2	Hologram	168
4.1.3	Reconstruction	170
4.2	Holographic Interferometry	172
4.2.1	Interference Fringes in a Double Exposure Hologram	173
4.2.2	Interference Fringes in a Real Time Hologram	175
4.3	Description by Means of Ellipses	176
4.4	Other Arrangements	178
4.5	Displacement Vector Lying on Plane <i>ACB</i>	179
4.5.1	Case Study of Displacements Determination in a Transparent Model	179
4.6	Displacement Vector Not Lying on Plane <i>ACB</i>	180
4.6.1	Order Determination When No Zero Fringe Is Present.	183
4.7	Strains from Displacements	184
	References.	185
5	Overview of Digital Image Correlation	187
5.1	Introduction	187
5.2	Operating Principle	189
5.3	Computation Area (Facet)	190
5.4	The Speckle Pattern	191
5.4.1	Case Study: Optimization of the Speckle Pattern.	194
5.5	Operating Principle of 2D Image Correlation	195
5.5.1	Displacements and Strains	196
5.6	3D Image Correlation	199
5.6.1	Stereoscopic Vision.	199
5.6.2	Rectification.	200
5.6.3	Stereoscopic Matching.	201
5.6.4	3D Correlation	201
5.7	Typical Arrangement of a DIC System	201
5.8	Theoretical and Practical Problems with DIC	204
5.8.1	Possible Problems Associated with Stereoscopic Matching	204
5.8.2	Observation Angle	205
5.8.3	Noise of the Input Digital Images	205
5.8.4	To Filter, or Not to Filter?	207
5.9	Case Study: Optimization of the Hardware and Software Parameters	209

5.10 Case Study: Strain Distribution in the Human Tibia 210
 5.11 Digital Volume Correlation 210
 References. 212

Part II Developing Phenomenological Theories and Problem Solving

6 Static Stress Models 217
 6.1 The Illustrative Advantage of a Full Field Analysis. 217
 6.1.1 Force at a Point of a Straight Boundary. 217
 6.1.2 Hertzian Contact. 220
 6.2 Design by Experiments of Pressure Vessels 224
 6.2.1 Case Study of Steam Generator Reactor
 Head with Four Openings on the Spherical Dome. 225
 6.2.2 Case Study of a Reactor Pressure Vessel
 with Four Nozzles in the Cylindrical Part 225
 6.3 Stress Concentration Factors. 227
 6.3.1 Basic Theoretical Cases: Stress Concentration
 for Circular Holes in Thin Plates 229
 6.3.2 Elliptic Hole. 239
 6.3.3 Concentration Factors for Slender Beams. 240
 6.4 Stress Beyond the Elastic Limit 241
 6.4.1 Imposed Displacements and Imposed Forces 241
 6.4.2 Models for the Simulation of Constitutive
 Materials Laws. 244
 6.4.3 Experimental Verification of the Neuber Model 246
 6.4.4 How Stress Modifies Material Behavior. 248
 6.4.5 How Material Behavior Modifies Stress State. 249
 6.4.6 Elementary Models of Physical Theory of Fracture 251
 6.5 Special Equipment for Static Tests 254
 References. 256

7 Local Stress Models for Variable Loads. 259
 7.1 Stress Analysis for Developing a Theory on High
 Cycles Fatigue 259
 7.2 Identification of Variable Loads 260
 7.2.1 Case Study of Fatigue Loads
 on a Sport Equipment 263
 7.2.2 Types of Loads 264
 7.3 Laboratory Testing 265
 7.3.1 Rotating Bending Test. 266
 7.3.2 Tension/Compression Test and Bending
 Test in One Plane. 267
 7.3.3 Torsion Test. 268

7.4	Fatigue Data Processing	269
7.4.1	Fatigue Data Collection	269
7.4.2	Model for Fatigue Behavior	271
7.5	Models for Notch Effect	273
7.5.1	Models for the Micro-plasticity Zone.	274
7.6	Support Factor	275
7.6.1	Support Factor in Slender Bodies	276
7.7	From Slender to 3D-Bodies	277
7.7.1	Modeling Non-uniform Stress States by Means of the Relative Stress Gradient	278
7.7.2	Relative Stress Gradient for Three Dimensional Bodies	282
7.8	Dependence of Fatigue Strength on the Support Factor	282
7.8.1	Size Effect	283
7.8.2	Model of the Support Factor as a Function of the Relative Stress Gradient	284
7.8.3	Modeling Support Effect Through Rotating Bending Fatigue Tests: Case Study on Characterization of a Titanium Alloy	289
7.9	Models for the Evaluation of Mean Stress Effect	296
7.10	Fatigue Assessment of Mechanical Components Through Relative Stress Gradient Theory.	299
7.10.1	Case Study for the Fatigue Assessment at a Three-Dimensional Protrusion in an Excavator Arm.	300
7.11	Macroscopic Fractography as Stress Analysis Tool	303
7.11.1	Rotating Bending of a Shaft with Two Diameters	305
7.11.2	Shafts in Bending at a Stress Rate $R = 0$	307
7.11.3	Shafts in Bending at a Stress Rate $R = -1$	308
7.11.4	Shafts in Torsion at Shear Stress Rate $R = -1$	308
7.12	Influence of the Surface Finishing.	310
7.13	Dixon Method	310
7.13.1	Case Study on the Determination of the Local Fatigue Limit of a Mechanical Component	313
7.13.2	Statistical Analysis of the Results	319
	References.	321
8	Local Strain Models for Variable Loads	325
8.1	Experimental Analysis of Local Deformation	326
8.2	Laboratory Testing	326
8.3	Strain-Controlled Versus Stress-Controlled Tests.	328

8.3.1	Repeated Loading in Stress-Controlled Tests	329
8.3.2	Repeated Loading in Strain-Controlled Tests	329
8.4	Model for Material Behavior	331
8.4.1	Concentration Factors for Variable Loads	333
8.5	Fatigue Life Model in Strain Controlled Problems.	334
8.5.1	Influence of the Mean Stress	336
8.6	Model for Variable Amplitudes	336
8.6.1	Cycle Counting Method.	338
8.6.2	Damage Per Cycle and Accumulated Damage	338
8.6.3	Application of Neuber Hypothesis Without Concentration Factors	339
8.7	Case Study of a Damage Evaluation	341
8.8	Torsional Low Cycle Fatigue	343
8.8.1	Ramberg-Osgood Model for Torsional Loads	344
8.8.2	Neuber Model for Torsional Loads	346
8.8.3	Fatigue Curve for Torsional Loads	346
8.8.4	Effect of Mean Shear Stress	348
8.9	Case Study of Notched Shafts in Torsion.	349
8.9.1	Special Equipment for Material Characterization in Shear Strain-Controlled Mode.	349
8.9.2	Determination of Cyclic Curve and Fatigue Curve Parameters	350
8.9.3	Special Equipments for Torsional Tests on Shafts	352
8.9.4	Torsional Tests with Assigned Loads Sequence	353
8.9.5	Appearance of the Cracks Directions in Torsional Low Cycle Fatigue	354
8.10	Preventing Loads Misalignments.	356
8.10.1	Functional Analysis of the Fixtures	357
8.10.2	Adopted Solution	358
	References.	360
9	Elementary Models for Stress Singularities	365
9.1	Behavior Classification	365
9.1.1	Elastic Behavior	366
9.1.2	Crack Loading Modes	367
9.1.3	Plane Stress and Plane Strain	367
9.2	The Field Equations for the Region Close to the Crack Tip	368
9.2.1	The Westergaard Stress Function	369
9.2.2	Displacements Close to the Crack Tip.	374
9.2.3	Stress Field for Modes II and III.	375
9.2.4	Magnification Factor	376
9.2.5	More Terms of Series Expansion	377
9.3	Identification of the Stress Intensity Factor Utilizing More Terms of the Series.	378

- 9.4 Crack Tip Plasticity. 380
 - 9.4.1 Apparent Plastic Zone Size According to Von Mises Criterion. 380
 - 9.4.2 Apparent Plastic Zone Size According to the Tresca Criterion. 382
 - 9.4.3 Apparent and True Plastic Zone Size. 384
- 9.5 Case Studies of the Stress Intensity Factor Identification for Internal Cracks 385
 - 9.5.1 Identification of SIF by Photoelasticity 386
 - 9.5.2 Identification of SIF by Holographic Interferometry. 397
 - 9.5.3 Identification of the Stress Intensity Factor by Strain Gages 400
- 9.6 Stress Analysis for Brittle Fracture 403
 - 9.6.1 Experimental Equipments. 404
 - 9.6.2 Fracture Mechanics Testing 409
 - 9.6.3 A Case Study on the Temperature Influence on SIF. 412
- 9.7 Models for Sub-critical Fatigue Crack Growth 414
 - 9.7.1 Case Study on Sub-critical Crack Growth for Turbine Steels and Design Criteria. 417
- 9.8 Visual Analysis of Crack Propagation 420
- References. 422

- 10 Stress Models in Biomechanics 425**
 - 10.1 Introduction 425
 - 10.1.1 Biomechanics of the Musculoskeletal System. 426
 - 10.1.2 Methods for Measuring Musculoskeletal Kinematics and Dynamics 427
 - 10.1.3 Biomechanical Musculoskeletal Models. 429
 - 10.2 Biomechanical Investigations in Orthopaedics at Different Dimensional Scales 431
 - 10.3 Case Study: Strain Distribution in the Human Tibia, a Uniform Stress Structure 436
 - 10.4 Case Study: Bone Strains Caused by Press Fitting of a Hip Stem 437
 - 10.5 Case Study: Deformations of an Acetabular Cup by Means of Holographic Interferometry 438
 - 10.6 Biomechanical Risk Assessment: Experimental Stress Analysis and Modelling 439
 - 10.6.1 Risk Analysis. 440
 - 10.6.2 Case Study: The Paradigm for Pre-clinical Validation of New Implantable Devices. 442
 - References. 445

- 11 Reliability Models Based on Experiments.** 449
 - 11.1 Introduction 449
 - 11.2 Determination of the Distributions of Sets of Data 451
 - 11.2.1 Case Study: Height Distribution in a Class of Students. 452
 - 11.2.2 Case Study: Probability of Static Failure of a Tensioned Rod. 457
 - 11.2.3 Case Study: Quality Control of Pin Diameters 461
 - 11.3 How to Linearly Interpolate a Set of Data and How to Account for Their Scatter. 465
 - 11.3.1 How to Determine the Coefficient of Plasticity and the Hardening Exponent of a Material and Related Distributions. 468
 - 11.3.2 How to Determine the *Manson-Coffin* Curve Parameters and Related Distributions. 471
 - 11.4 Multi-variable Problems: Application of a *Most Probable Point Method* 475
 - 11.4.1 Low Cycle Fatigue on Turbogenerators and Probability of Failure 482
 - 11.4.2 Case Study: Probability of Failure of a Turbogenerator Rotor 485
 - 11.4.3 Fracture Mechanics and Reliability Assessment 488
 - 11.4.4 Case Study: Fracture Mechanics and Reliability Assessment 490
 - References. 492
- Index** 495

Introduction

The main aim of the book is the discussion of Experimental Stress Analysis as a methodology to support theoretical models and build phenomenological theories.

After a preliminary chapter that introduces the concept of inverse problems and tackles the methods used for solving them (Chap. 1), the first part of the book briefly summarizes the conceptual basis of the experimental stress analysis methods utilized by the authors, which are judged to be useful either for illustration or practical use (Chaps. 2–5). The description is thus limited to a classic presentation—sometimes highlighting aspects that are not always detailed in the traditional literature—since it is considered more important to demonstrate their usefulness for clarifying practical engineering problems than to provide information about the latest developments of the methods.

The chapters in the second part of the book are devoted to the application of stress analysis to several case studies of technical interest, demonstrating the conceptual and practical importance of the inverse approach (Chaps. 6–9). In this part of the book we deemed it appropriate to summarize some results of the classic analytical theories which, although essential in academic training, generally have no place in books of this type, such as the theory of elasticity and the elementary stress theory of Fracture Mechanics. The presentation is limited to a few but important theories of modern engineering.¹ Chapter 10 is devoted to the application of experimental stress analysis to biomechanical systems and offers a vivid example of reliability-related engineering problems. Finally, the last chapter describes how to

¹ A review of *Hitachi* from the 1970s (but likely to be still valid today) selected the following items as the main reasons for failures with the corresponding incidences on total failures and number of artifacts of its production:

- fatigue (59,8 %)
- static fracture and creep (13,4 %)
- brittle fracture (10,3 %)
- buckling and excess of deformation (8,9 %)
- wear and fretting (4,0 %)
- corrosion (3,6 %).

utilize experimental stress analysis for reliability conclusions, with reference to a case study that deals with the life extension of turbo-machinery (Chap. 11).

The book organization is summarized in the matrix of Fig. 1, where the rows represent the methods and the columns the structural mechanics fields, while the applications are located at the intersections.

		Applications	Statics	HC Fatigue	LC Fatigue	Fracture Mech	Biomechanics
Inverse Approach	Methods						
	Strain gauges		↓	↓	↓	↓	↓
	Photoelasticity			↓		↓	↓
	Holographic Interferometry					↓	↓
	DIC						↓
	Reliability Assessment		↓	↓	↓	↓	↓

Fig. 1 Book organization

Instructions to Instructors

Following these premises, the book can offer materials for two courses at Engineering Schools: the first, of about 60 hours, aimed at undergraduate students and covering a basic syllabus on Experimental Stress Analysis methods and application to elementary theory of elasticity. Active exercises for groups of no more than six-seven students should be devised on the following topics:

- Strain gages mounting and connection to circuits.
- Load-cell building and calibration.
- Optical analysis of field stress.
- Comparison between analytical and experimental results.

A second course is suitable for graduate students. The course syllabus, corresponding to approximately 60 hours, should be oriented to the applications of the methods to the main topics of structural mechanics, focusing on items developed in the laboratory, such as:

- Strain gages measurements in actual mechanical parts.
- Static and fatigue tests performed on universal and special testing machines.
- Fracture Mechanics tests on specimens.
- Illustrations in a Biomechanics Laboratory of tests on bones and prostheses by strain gages, photoelastic coating, digital image correlation.
- The course can be concluded by a Reliability treatment of experimental data.

Part I

A Brief Review of the Experimental Methods Utilized in the Book

The first part of the book briefly describes some common electrical (bonded strain gages) and optical (photoelasticity, holography and digital image correlation) methods and their relative merits and limitations, as they are used by the authors for application to stress analysis. Specialized texts are the best further reference for readers interested in each of them.

The chapters review the basic formulas and essential concepts, looking, however, at some specific questions that are not always emphasized in classical books, and avoiding unnecessarily complex experimental equipments. Indeed, many techniques can be applied in a relatively simple way without losing the great benefit of *understanding the problem* before searching for a solution for it. In other words, understanding the main variables of the phenomenon under study and clarifying their effects is more important than finding a solution to a single case.

Chapter 1

Introduction to Inverse Problems

Abstract Experimental Stress Analysis has been traditionally applied—through a *direct or forward* approach—for solving structural mechanical problems as an alternative and complementary methodology to the theoretical one. The great development of numerical methods has largely overruled this task. In addition, the increased accuracy of numerical tools has confined the forward approach to the role of experimental verification restricted to cases of complex and non-conventional numerical modeling, such as stress states resulting from singularities, material anisotropy, etc. If, however, causes (such as forces, impressed temperatures, imposed deformations) or system parameters such as geometry, materials and boundary conditions are unknown, the case is totally different, and the experimental *inverse* approach has no alternatives. Through measurements of the effects like displacements, strains and stresses, it is possible to find solutions to these inverse problems by identifying the unknown causes, integrating a series of experimental data into a theoretical model. The accuracy of data together with a proper selection of the quantities that must be measured are a necessary premise for limiting the experimental errors that can influence the accuracy of the inversely estimated results.

1.1 Premise

The inverse problem consists in using the results of actual measurements to infer the values of the parameters characterizing the system under investigation. Measurements are justified when structural systems or physical models are available and when forces or system parameters as stiffness or a function of these variables are unknown and must be identified.

Practical recent and promising applications of the inverse problems can be found in activities of supervising and health-monitoring existing structures [1], controlling their deterioration and estimating their residual life. This is one of the most relevant tasks in modern societies, sensitive to ecology and sustainability of the technical solutions.

The present chapter must be considered a primer of this approach that totally has reoriented the experimental stress analysis, better defining its main task and its

fundamental benefit. For a complete presentation of inverse methods see e.g., the books of [2–6] and the relative bibliography.

1.1.1 General Rules

All problems in mathematics, physics and engineering are divided into two classes: well-posed problems and ill-posed problems.

Direct problems can be ill-posed when they are not well defined and inverse ones can be ill-posed when the experimental errors (noise) are relevant; considering here the inverse problems, it could happen that [2]:

Some of them are well-posed problems. For these:

- A solution exists.
- The solution is unique.
- A small change of data leads to a small change in the solution.

Vice-versa, a broad number of inverse problems belongs to the class of ill-posed problems. For these:

- A solution may not exist.
- There may be more than one solution.
- Small variability of experimental data may originate great variability on the estimation of unknown variables.

1.1.2 Rules for Inverse Problems

The approach to inverse problems needs the developing of the following steps [5]:

- Define the unknowns to be identified.
- Arrange a forward theoretical or numerical model that must be representative of the physical problem, to fit to measured data. The solution of this process is called *Identification* of unknown variables.
- Estimate whether the system can be theoretically formulated or not in an explicit matrix form. The algebraic relationships in matrix form are suitable only for structural discretized systems.

In the present chapter the first case studies are introduced in discretized form, and only in other chapters will some cases be presented for non-discretized systems.

- Develop an preliminary Design of Experiment on the forward problem for analyzing the sensitivity of the factors that must be independently influential to the outputs (effects), trying to avoid the combined influence of several variables, i.e. interaction effects.
- Estimate the level of experimental errors on the effects (i.e. signal to noise ratio) and, if it is possible, select effects less influenced by experimental errors. The experimental errors can favor the *ill-posedness* of the problem.

- If the system can be formulated with matrix form and the number of known variables equals the number of unknown variables, the problem is determined or even-posed and inversion of the system matrix (or transformation, or operator matrix) can be performed through a classical algorithm to obtain the inverse solution.

When the number of known and unknown variables is not the same, matrix inversion methods can be used, all of them leading to approximate solutions. The pseudo-inverse matrix concept must be introduced.

- Utilize regularization techniques for improve ill-posed problems. Some examples in Chap. 2 will clarify their use and effectiveness.
- If the system cannot be formulated with matrix form, the problem must be dealt by formulation of a *functional of error* between forward and inverse solutions. It must be minimized through optimization/minimization techniques. This procedure provides the unknown parameters of the forward model.
- Verify the solution, introducing the identified variables in the model, comparing the theoretical results with the experimental data.

1.2 Forward and Inverse Problems for Elastic Discretized Structures

The germinal idea of the elastic structural analysis is in the Hooke’s Law *Ut Tensio, sic Vis* (as the extension, so the force) that, for many variables, is applied to discrete systems or to continuous systems discretized through a representation in Finite Elements. The law for the forward problem is written with the formalism of Eq. 1.1, Fig. 1.1.

$$\mathbf{y} = \mathbf{A}(\mathbf{p}) \cdot \mathbf{x} \tag{1.1}$$

with these symbols meaning:

- \mathbf{y} output displacements vector
- \mathbf{x} input forces vector

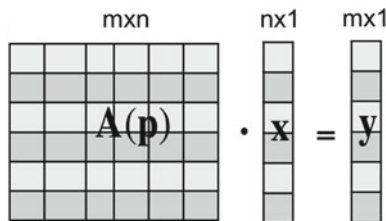


Fig. 1.1 Matrix relationship for structural problems. For one variable x and y are numbers, for several variables \mathbf{x} and \mathbf{y} are vectors. If \mathbf{x} and \mathbf{y} have the same dimensions the matrix is square

$\mathbf{A} = \mathbf{A}(\mathbf{p})$ system matrix (or transformation matrix), function of \mathbf{p} , a vector of all data necessary for defining geometry, material properties and boundary conditions.

\mathbf{A}^{-1} inverse of the system matrix

If \mathbf{A} is a positive definite square ($m \times m$) matrix the problem is well-posed and an unequivocal solution exists:

$$\mathbf{x} = \mathbf{A}^{-1} \cdot \mathbf{y} \quad (1.2)$$

Given a geometrically complex system, the forward problem, from the known forces to the unknown displacements, needs the use of numerical methods that schematized (meshed) the structure through a great number of finite elements.

It is clear that the problem is theoretically reversible and, in this case, it is only formal to take the Eq. 1.1 as a forward problem and the Eq. 1.2 as an inverse problem. In the Finite Element formulation, the Eq. 1.2 is chosen as a forward solver, because the vector \mathbf{x} and the system matrix \mathbf{A} are given and the displacements vector \mathbf{y} is obtained by the matrix inversion.

Since in the forward problems it is recommended to have the same number of known and unknown variables, a unique solution is found, solving a set of linear equations, by common solution codes. The modern methods do not necessarily need the formal inversion of the matrix. Nevertheless, the solution is unequivocal. Figure 1.2 shows e.g., the simulation of a detail of a crane arm [7, 8] with a great number of tetrahedral finite elements.

Having, instead of the Eq. 1.2, a mathematical forward solver written in opposite way:

$$\mathbf{x} = \tilde{\mathbf{A}}(\mathbf{p}) \cdot \mathbf{y} \quad (1.3)$$

if $\tilde{\mathbf{A}}$ and \mathbf{y} are known a direct problem is given, while if \mathbf{A} and \mathbf{x} are known, an inverse problem is given, Fig. 1.3. The Table 1.1 synthesizes the different types of forward and inverse problems (from Laermann [4]).



Fig. 1.2 The forward numerical model as a reference model for experimental measurements. The finite elements method reduces complex structures to a numerous series of simple elements for which the system matrix is assembled. See a case study in Sect. 7.10.1 in Chap. 7

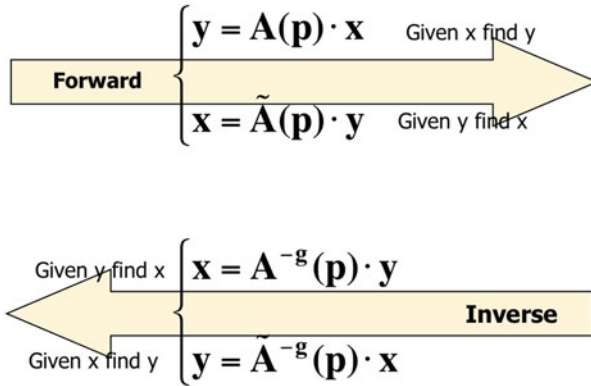


Fig. 1.3 Scheme of forward and inverse problems

Table 1.1 Different types of forward and inverse problems

Case	Forward solver	Given	Obtained	Type
–		\mathbf{p}, \mathbf{x} or \mathbf{p}, \mathbf{y}	\mathbf{y} or \mathbf{x}	Forward
I	$\mathbf{y} = \mathbf{A}(\mathbf{p}) \cdot \mathbf{x}$ or $\mathbf{x} = \mathbf{A}(\mathbf{p}) \cdot \mathbf{x}$	\mathbf{p}, \mathbf{y} or \mathbf{p}, \mathbf{x}	\mathbf{x} or \mathbf{y}	1st kind
II		\mathbf{x} or \mathbf{y}	\mathbf{p}	2nd kind
III ^a	$\mathbf{y}_g, \mathbf{y}_u = \mathbf{A}(\mathbf{p}_g, \mathbf{p}_u) \cdot (\mathbf{x}_g, \mathbf{x}_u)$	$\mathbf{p}_g, \mathbf{x}_g, \mathbf{y}_g$	$\mathbf{p}_u, \mathbf{x}_u, \mathbf{y}_u$	Mixed

Taken from Laermann [4]

^a Subscript g is for given quantities, subscript u for unknown variables

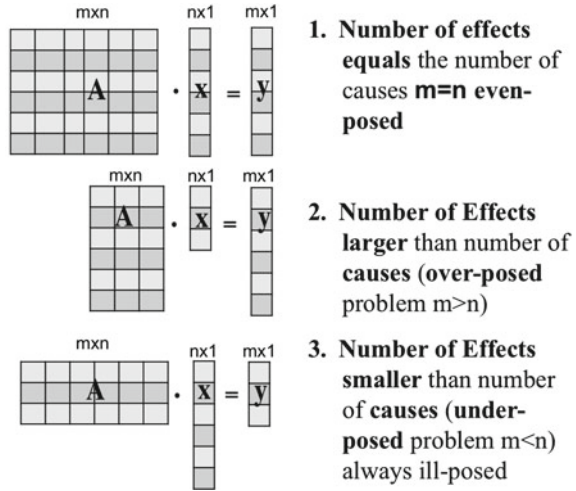
The causes can be external loads, imposed temperatures, imposed displacements, while the unknown variables that designate the effects, are quantities that can be experimentally measured as stresses, strains, natural frequencies, etc.

Only rarely is the number of the known variables equal to the number of unknown variables. Given e.g. \mathbf{x} the vector of forces applied to the system \mathbf{A} , if \mathbf{y} is the displacements vector, the following alternatives might happen:

- If displacements \mathbf{y} are measured (output) and the system parameters \mathbf{p} are known, the unknown input forces \mathbf{x} can be identified.
Or:
- If displacements \mathbf{y} are measured and the forces \mathbf{x} are known, the system parameters of the transformation matrix \mathbf{A} can be identified.
- If all the other parameters of the system are given and the inputs and outputs of the system are known, the boundary conditions that are a combination of forces and displacements can be identified.
- When the number of input variables is the same for output variables, the matrix of the forward problem is square and the problem is determined,¹ and the

¹ Called also even-posed.

Fig. 1.4 Different kinds of matrices when the number of known variables (input) is equal to, less or greater than the number of unknown variables (output)



identification is done without approximation. Nevertheless, also in this case, the problem can be well-posed or ill-posed, depending on the role of the experimental errors.

- When the number of input variables is less than the number of output variables, the matrix of the forward problem is rectangular and the problem is over-determined (called also over-posed) and identification is possible but only approximately. The problem can be well-posed or ill-posed.
- When the number of unknown variables is greater than the number of measured variables, the matrix of the forward problem is rectangular and the problem is under-determined (called also under-posed) and always ill-posed.

The different kinds of matrices of the examined cases are shown in Fig. 1.4.

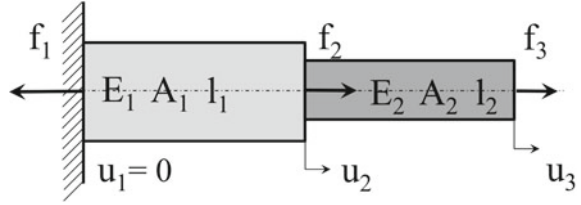
Over-posed and under-posed problems can give only approximate solutions because the system matrices are not square, and rigorous inverted matrices A^{-1} do not exist. Approximate solutions are given by the pseudo-inversion of the system matrices.

Summarizing, to solve an inverse problem a suitable theoretical (or numerical) model and a set of accurate experimental data are necessary. The accuracy and a proper selection of measured variables are necessary for limiting the experimental errors that are transmitted by the inversion process to the unknown variables, with the possibility of causing ill-posed solutions.

Let us report some Exercises taken from Liu and Han [5] that clarify the meaning of direct and of inverse problems.

Exercise 1.1 (*Inverse Problems for Uniaxial Load. First Case: Determined or Even-posed*) Consider a structural system with two bars of different elastic modulus,

Fig. 1.5 Model of a composite bar



transversal section and length, uniaxially loaded and constrained as shown in Fig. 1.5.² The stiffness parameters of the two parts are respectively $k_1 = E_1 A_1 / l_1$ and $k_2 = E_2 A_2 / l_2$.

(a) Forces, material properties and geometry are known and the *theoretical solution* of the forward problem is:

$$\begin{cases} u_2 = \frac{f_2}{k_1} + \frac{f_3}{k_1} \\ u_3 - u_2 = \frac{f_3}{k_2} \end{cases} \quad (1.4)$$

or:

$$\begin{cases} u_2 = \frac{f_2}{k_1} + \frac{f_3}{k_1} \\ u_3 = u_2 + \frac{f_3}{k_2} = \frac{f_2}{k_1} + f_3 \left(\frac{1}{k_1} + \frac{1}{k_2} \right) \end{cases}$$

and in matrix form:

$$\begin{pmatrix} u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{k_1} & \frac{1}{k_1} \\ \frac{1}{k_1} & \frac{1}{k_1} + \frac{1}{k_2} \end{pmatrix} \cdot \begin{pmatrix} f_2 \\ f_3 \end{pmatrix} \quad (1.5)$$

or:

$$\mathbf{y} = \mathbf{A}(k_1, k_2) \cdot \mathbf{x}$$

The solution of the relative inverse problem is obtained by the inversion of the system matrix, (see e.g. algorithm *Inverse* from Mathematica[®] in [9]).

Forces Identification

Thus the forces can be identified by the displacements through the equation:

$$\begin{pmatrix} f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 \end{pmatrix} \cdot \begin{pmatrix} u_2 \\ u_3 \end{pmatrix} \quad (1.6)$$

² Example taken from Liu and Han [5].

or:

$$\begin{cases} f_2 = k_1 u_2 - k_2(u_3 - u_2) \\ f_3 = k_2(u_3 - u_2) \end{cases} \quad (1.7)$$

or in matrix form:

$$\mathbf{x} = \mathbf{A}^{-1} \cdot \mathbf{y}$$

Geometric and Elastic Parameters Identification

From the same Eq. 1.6 is possible to obtain the stiffnesses as functions of the displacements:

$$\begin{pmatrix} k_1 \\ k_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{u_2} & \frac{1}{u_2} \\ 0 & \frac{1}{u_3 - u_2} \end{pmatrix} \cdot \begin{pmatrix} f_2 \\ f_3 \end{pmatrix} \quad (1.8)$$

From Eq. 1.8, the stiffness or the Young's modulus or the lengths of both bars members can be easily determined, remembering that $k = EA/l$.

The forward problem can be ill-posed for a lack of boundary conditions or in the case of k_1 or k_2 equal to zero, Eq. 1.5, while the inverse problem can be ill-posed if u_2 and u_3 , (that are small in elastic loaded structures) are also close to one another. In these cases the error on estimated forces or stiffnesses can become relevant, Eqs. 1.7 and 1.8.

Exercise 1.2 (*Inverse Problems for Uniaxial Load. Second Case: Over-determined or over-posed*) Determine the stiffnesses of the two bars, supposing that $k_1 = k_2 = k$.

(a) Equations 1.5 and 1.6 give [5]:

$$\begin{pmatrix} u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{k} & \frac{1}{k} \\ \frac{1}{k} & \frac{2}{k} \end{pmatrix} \cdot \begin{pmatrix} f_2 \\ f_3 \end{pmatrix} \quad (1.9)$$

$$\begin{pmatrix} f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} 2k & -k \\ -k & k \end{pmatrix} \cdot \begin{pmatrix} u_2 \\ u_3 \end{pmatrix} \quad (1.10)$$

This assumption does not influence the identification of the forces but over-determines the stiffnesses of the two parts of the bar that can be identified in two different ways and the values can be different (due to experimental errors):

$$\begin{cases} k = \frac{f_2}{2u_2 - u_3} \\ k = \frac{f_3}{u_3 - u_2} \end{cases}$$

(b) The same difficulty arises for the identification of the elastic modulus, transverse section areas and lengths, (if assumed identical in both bars), thus originating an *over-determined* inverse problem. Also in this case the problem can be ill-posed if the two displacements are affected by significant experimental errors.

This case, extended to many data, is the most common in laboratory experimentation, because the experiment is generally planned in such a way to have redundancy of data.

Recourse is then done to a method of minimization of the differences and the most common is the *Least Squares Method* (LSM), for finding a reasonable approximate mean value that is assumed as the most probable. In the following Chaps. 2 and 9, examples of this minimization techniques will be shown for strain gages application in multi-axial loading cells as well as in Fracture Mechanics for identification of Stress Intensity Factors. In the last case it will no longer be possible to have a reference to a discrete system and a general optimization problem for linear and non linear systems will be adopted.

Exercise 1.3 (*Inverse Problems for Uniaxial Load. Third Case: Under-determined, then ill-posed*) This is the case of a lack of experimental data with respect to the number of unknowns to be determined.

With reference to Fig. 1.5, if e.g. only u_3 is measured, the second of Eq. 1.9 gives [5]:

$$u_3 = \frac{f_2}{k_1} + f_3 \frac{k_1 + k_2}{k_1 k_2} \quad (1.11)$$

that is the equation of a straight line:

$$a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 = \mathbf{y}^{measured} \quad (1.12)$$

Infinite solutions for pairs of values $x_2 = f_2$ and $x_3 = f_3$ exist, i.e. the solution is non-unique. The problem is always ill-posed.

Exercise 1.4 (*Inverse Problems for Uniaxial Load. Fourth Case: Even-posed for Boundary Condition Identification*) In this case everything but the boundary conditions is known. Instead of the first of Eq. 1.4 here repeated:

$$\begin{cases} u_2 = \frac{f_2}{k_1} + \frac{f_3}{k_1} \\ u_3 - u_2 = \frac{f_3}{k_2} \end{cases} \quad (1.13)$$

this new equation is written:

$$\begin{cases} u_2 - u_1 = \frac{f_2}{k_1} + \frac{f_3}{k_1} \\ u_3 - u_2 = \frac{f_3}{k_2} \end{cases}$$

or

$$\begin{pmatrix} -1 & k_1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} f_1 \\ u_1 \end{pmatrix} = \begin{pmatrix} k_1 u_2 \\ f_2 + k_2(u_3 - u_2) \end{pmatrix} \quad (1.14)$$

Thus, the unknown boundary conditions are identified:

$$\begin{pmatrix} f_1 \\ u_1 \end{pmatrix} = \begin{pmatrix} -1 & k_1 \\ 1 & 0 \end{pmatrix}^{-1} \cdot \begin{pmatrix} k_1 u_2 \\ f_2 + k_2(u_3 - u_2) \end{pmatrix} \quad (1.15)$$

Solving Eq. 1.15 for the unknown input of boundary values, gives:

$$\begin{pmatrix} f_1 \\ u_1 \end{pmatrix} = \begin{pmatrix} f_2 + k_2(u_3 - u_2) \\ u_2 + \frac{f_2 + k_2(u_3 - u_2)}{k_1} \end{pmatrix} \quad (1.16)$$

The number of unknowns is the same as the number of knowns, thus this problem is determined or even-posed. The input and the output vectors consist of both components of displacement and force. The boundary conditions can be determined if all the other system parameters are given and the causes and the effects are known.

It was remarked that, as long as we treat discrete structural problems described by matrix algebra, it may be argued that forward and inverse problems are mathematically reciprocal. In spite of this, the influence of the experimental errors on the unknown identification was shown. For continuous problems with differential or integral operators the problem of error transmission is exacerbated. Leaving this aspect to specialized books, we must now reflect upon the quality of the predication after the inversion, and suitable indexes must be introduced to quantify it.

1.2.1 Quality Indicators of Inverse Solutions

The pseudo-inversion of a matrix is an approximate operation: it means that the product of this pseudo-inverse and of the original matrix do not necessarily give the identity matrix.

There are indicators of the quality of the solution that are necessarily verified, every time a matrix inversion is performed. They are called:

1. *Data or output* resolution or output reproducibility matrix.
2. *Model or input* resolution or input reproducibility matrix.

Writing again Eq. 1.1:

$$\mathbf{y} = \mathbf{A} \cdot \mathbf{x} \quad (1.17)$$

in the case of a not square matrix, the real inverse does not exist and the estimate of the solution \mathbf{x} assumes the form:

$$\mathbf{x}^{estim} = \mathbf{A}^{-g} \cdot \mathbf{y}^{measur} \quad (1.18)$$

where: \mathbf{A}^{-g} is the pseudo-inverse matrix, according to the scheme shown in Fig. 1.3.

There is no reason that the product $\mathbf{A}^{-g} \cdot \mathbf{A}$ has to be necessarily equal to identity matrix \mathbf{I} , but only this value, referring to the unitary matrix, is an index of the goodness of the inversion. The effect-vector, with estimated value of the cause, is:

$$\mathbf{y}^{predict} = \mathbf{A} \cdot \mathbf{x}^{estim} \quad (1.19)$$

Substituting Eq. 1.18 in Eq. 1.19 gives:

$$\mathbf{y}^{predict} = \mathbf{A} \cdot \mathbf{A}^{-g} \cdot \mathbf{y}^{meas} = \mathbf{D} \cdot \mathbf{y}^{meas} \quad (1.20)$$

This equation shows that, if $\mathbf{A}^{-g} \cdot \mathbf{A} = \mathbf{I}$ where \mathbf{I} is an identity matrix, the inverse procedure *reproduces* the measurement data. If $\mathbf{D} \neq \mathbf{I}$, the measurement data will not be reproduced (or resolved). Therefore the matrix \mathbf{D} is defined as an *data reproducibility or data resolution matrix*.

Furthermore, if it is argued that the true values \mathbf{x}^{true} should satisfy the forward model, substituting the \mathbf{x}^{true} in Eq. 1.19 the real measured values should be obtained: $\mathbf{y}^{measur} = \mathbf{A} \cdot \mathbf{x}^{true}$. Substituting this relationship in Eq. 1.18

$$\mathbf{x}^{estim} = \mathbf{A}^{-g} \cdot \mathbf{y}^{measur} = \mathbf{A}^{-g} \cdot \mathbf{A} \cdot \mathbf{x}^{true} = \mathbf{G} \cdot \mathbf{x}^{true}$$

gives the matrix \mathbf{G} . If $\mathbf{G} \neq \mathbf{I}$, the estimation will not be the true input. Therefore this matrix is defined as an *input resolution or input reproducibility matrix*.

1.2.2 Inverse Solution for Systems in Matrix Form

As it was seen, if the system can be formulated in an explicit matrix form, a general inversion of the transformation matrix can be performed to obtain inverse or pseudo-inverse solutions.

1.2.2.1 Pseudo-Inverse Matrix for Over-Determined (or Over-Posed) ($m > n$) Problems

That is a very common case in the laboratory practice and in numerous applications. From the general equation

$$\mathbf{y} = \mathbf{A} \cdot \mathbf{x}$$

In the case of experimental analysis the following equation holds:

$$\mathbf{y}^{meas} - \mathbf{A} \cdot \mathbf{x} = \varepsilon$$

with ε = experimental error, $\mathbf{A} \in \mathbf{R}^{m \times n}$ and $m > n$.

The determination of $x \in \mathbf{R}^n$ is done under the condition of the minimization of the norm of the vector of the residuals that is the *distance* of the prediction of the results of the forward model from the measurements:

$$J = \|\mathbf{y}^{meas} - \mathbf{A} \cdot \mathbf{x}\|^2 = \min \quad (1.21)$$

From the minimum condition:

$$\frac{\partial J}{\partial \mathbf{x}} = 0 - \mathbf{A} \mathbf{y}^T (\mathbf{y}^{meas} - \mathbf{A} \cdot \mathbf{x}) = 0$$

the set of normal equations is derived:

$$\mathbf{A}^T \mathbf{y}^{meas} = \mathbf{A}^T \mathbf{A} \mathbf{x}^{estim} \quad (1.22)$$

Note that the matrix $\mathbf{A}^T \mathbf{A}$ is surely symmetric. Assuming it is also invertible:

$$\mathbf{x}^{estim} = \mathbf{A}^{-g} \cdot \mathbf{y}^{meas} \quad (1.23)$$

with the following expression for the pseudo-inverse matrix, Fig. 1.6:

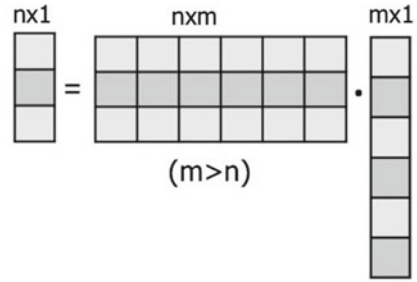
$$\mathbf{A}^{-g} = (\mathbf{A}^T \mathbf{A})^{-1} \cdot \mathbf{A}^T$$

The output resolution matrix is:

$$\mathbf{A} \cdot \mathbf{A}^{-g}$$

that is not an identity matrix. Vice-versa the input resolution matrix is:

Fig. 1.6 Solution for the inverse over-determined, thus over-posed problem



$$\mathbf{A}^{-g} \cdot \mathbf{A} = \mathbf{I} \tag{1.24}$$

This means the Least Squares Solution gives the true estimation of the model.

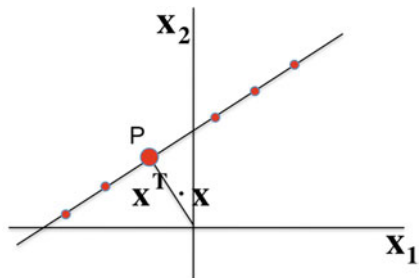
It could happen that the matrix $\mathbf{A}^T \mathbf{A}$ does not have a full rank, then it is not invertible. In such cases, it can become necessary to introduce some kind of regularization method, even accepting relatively approximate solutions.

1.2.2.2 Pseudo-Inverse Matrix for Under-Posed Problems ($m < n$)

It is the most critical inverse case that often arises in problems of maintenance, damage analysis and control of structural systems, and for the evaluation of residual life of structural systems. In these cases, it could be very difficult to have a sufficient number of measurements necessary to identify all the unknown parameters of the system.

An approximate result can be found if other information on the solution exists [4]. Without other conditions the problem offers infinite solutions that exactly satisfy the equation of the corresponding forward model without any error and can become determined only if it is possible to add other physically meaningful information. This is clarified by the simple mathematical model of a straight line. All the points of it satisfy the Eq. 1.12 and Fig. 1.7, thus infinite solutions exist.

Fig. 1.7 The minimum length solution for an under-posed problem



A formal method for having determined the problem, is the Minimum Length (ML) solution, but this solution is totally arbitrary if not supported by some physical evidence. Given the model:

$$\mathbf{y} = \mathbf{A} \cdot \mathbf{x}$$

it must be:

$$\mathbf{y}^{meas} - \mathbf{A} \cdot \mathbf{x} = 0$$

with $\mathbf{A} \in \mathbf{R}^{m \times n}$ and $m < n$.

Determine $\mathbf{x} \in \mathbf{R}^n$ such that:

$$J = \|\mathbf{y}^{meas} - \mathbf{A} \cdot \mathbf{x}\|^2 = \min \quad (1.25)$$

This is an undetermined problem and cannot be solved. According to the following Eq. 1.26

$$J = \mathbf{x}^T \cdot \mathbf{x} + \lambda^T \cdot (\mathbf{y}^{meas} - \mathbf{A}\mathbf{x}) = \min \quad (1.26)$$

The error functional is formed by two parts connected by a Lagrange multiplier: the first is the Pythagorean length of the vector \mathbf{x} and the second represents a constraint on the variable that forces the unknown vector \mathbf{x} to satisfy the equation system of the forward model. Therefore, the function J is, in fact, a constrained Pythagorean length of the unknown vector \mathbf{x} subjected at m constraints ($\mathbf{y}^{meas} - \mathbf{A} \cdot \mathbf{x} = 0$) for m Lagrangian multipliers.

To find the minimum length (that in Fig. 1.7 is represented by the distance from the origin), both conditions:

$$\frac{\partial J}{\partial \mathbf{x}} = 0$$

$$\frac{\partial J}{\partial \lambda} = 0$$

are required for writing the following equations:

$$\frac{\partial J}{\partial \mathbf{x}} = 2\mathbf{x} - \mathbf{A}^T \lambda = 0 \quad (1.27)$$

and:

$$\frac{\partial J}{\partial \lambda} = \mathbf{y}^{meas} - \mathbf{A} \cdot \mathbf{x} = 0 \quad (1.28)$$

Equation 1.18 gives:

$$\mathbf{x} = \frac{1}{2} \mathbf{A}^T \lambda \quad (1.29)$$

that must be substituted in Eq. 1.19:

$$2\mathbf{y}^{meas} = \mathbf{A}\mathbf{A}^T \cdot \lambda \quad (1.30)$$

If $\mathbf{A}\mathbf{A}^T$ is invertible, the vector of Lagrange multipliers λ is:

$$\lambda = 2(\mathbf{A} \cdot \mathbf{A}^T)^{-1} \mathbf{y}^{meas} \quad (1.31)$$

Substituting Eqs. 1.20 and 1.31 finally gives:

$$\mathbf{A}^{-g} = \mathbf{A}^T (\mathbf{A} \cdot \mathbf{A}^T)^{-1}. \quad (1.32)$$

1.3 Systems in Functional Form

For systems that cannot be formulated in an explicit matrix form, some methods exist for solving inverse problems, as reported in the bibliography of the cited books.

The general way consists in establishing a functional of residuals or errors defined with a suitable norm, and in utilizing optimization/minimization techniques to search for the solution that minimizes this functional.

Several minimization algorithms are offered by the mathematical literature and greatly simplify the researcher's task, such as the *FindFit* (Mathematica®) [9].³

In a case study on K parameter identification of Fracture Mechanics, (see Chap. 9), the forward problem cannot be expressed in matrix form, but only in functional forms of this type:

$$\mathbf{y} = \mathbf{A}(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k, \mathbf{x}) \quad (1.33)$$

where \mathbf{y} is a vector of the effects, \mathbf{A} is the system transformation matrix, function of vectors parameters \mathbf{p}_i , while \mathbf{x} is the inputs vector, identified by the solution of the inverse problem.

The error functional is the norm of a vector of the difference between the predicted output based on the forward model of Eq. 1.33 and the measured data that are presumed to be generated by the true value of input variables \mathbf{x}^{true} ; (n is the number of experimental data):

$$J(\mathbf{x}) = (\mathbf{y}^{pred} - \mathbf{y}^{meas})^T \cdot (\mathbf{y}^{pred} - \mathbf{y}^{meas}) = \sum_{i=1}^n [y_i^{pred}(\mathbf{x}) - y_i^{meas}(\mathbf{x}^{true})]^2 \quad (1.34)$$

³ E.g., the algorithm *FindFit* [*data*, *expr*, *pars*, *vars*] finds numerical values of the parameters *pars* that make *expr* give a best fit to data as a function of *vars*. The expression *expr* can depend either linearly or nonlinearly on the *par*_{*i*}. This algorithm is particularly useful for minimizing functional of error because it allows the use of several alternatives that can be select such as *Conjugate Gradient*, *Gradient*, *Levenberg Marquardt*, *Newton*, and *Quasi Newton* methods, with the default use of the *Least-Squares* methods.

If $\mathbf{x} \neq \mathbf{x}^{true}$:

$$J \geq 0$$

From this equation it is clear that if another type of functional is used, e.g. the L_1 norm that is the sum of absolute errors:

$$J = \sum_{i=1}^n |y_i^{pred}(\mathbf{x}) - y_i^{meas}(\mathbf{x}^{true})|$$

the values of the input variables that minimize this functional will be certainly different from the values obtained, e.g., from L_2 norm.⁴

1.3.1 Regularization Method of Tikhonov-Phillips

A method that generalizes the previous one and that is essential for solving under-posed problems but becomes useful also for over-determined problems (when they are ill-posed) is the Tikhonov-Phillips⁵ regularization method [4] that provides a way to make use of further information to regularize the solution. This method, also called *Damped Least Squares Solution*, defines the discrepancy functional in the following form:

$$J = (\mathbf{y}^{meas} - \mathbf{Ax})^T (\mathbf{y}^{meas} - \mathbf{Ax}) + \alpha \mathbf{x}^T \mathbf{x} = \min \quad (1.35)$$

or:

$$J = \|\mathbf{y}^{meas} - \mathbf{Ax}\|^2 + \alpha \|\mathbf{x}\|^2 = \min \quad (1.36)$$

The first term is the L_2 norm of the *Least Squares Solution* while the second term is the Pythagorean distance of the *Minimum Length Solution* and α denotes a regularization parameter called *damping factor* that is an additional unknown of the problem. Its task is to reduce the weight of the distance inside the functional, so that a too high α value improves the solution stability but reduces the solution accuracy and a too low value of the parameter improves the accuracy but does not modify the ill-posedness of the solution. Searching for the minimum error requires:

$$\frac{\partial J}{\partial \mathbf{x}} = 0$$

⁴ For each kind of engineering problem it is necessary to check which inverse solution method best meets the requirements of the solution. Different solution algorithms applied to the same data can yield different answers (see examples in Chap. 7).

⁵ Tikhonov regularization has been invented independently in many different contexts. Some authors use the term Tikhonov Phillips regularization.

that leads to:

$$2(\mathbf{y}^{meas} - \mathbf{Ax})(0 - \mathbf{A}) + 2\alpha\mathbf{xI} = 0 \quad (1.37)$$

$$-\mathbf{A}^T(\mathbf{y}^{meas} - \mathbf{Ax}) + \alpha\mathbf{xI} = 0 \quad (1.38)$$

then:

$$\mathbf{A}^T \mathbf{y}^{meas} = \mathbf{A}^T \mathbf{Ax} + \alpha\mathbf{xI} \quad (1.39)$$

Matrix $(\mathbf{A}^T \mathbf{A} + \alpha\mathbf{I})$ is symmetric and invertible. The solution is:

$$\mathbf{x}^{estim} = (\mathbf{A}^T \mathbf{A} + \alpha\mathbf{I})^{-1} \mathbf{A}^T \cdot \mathbf{y}^{meas} = \mathbf{A}^{-g} \cdot \mathbf{y}^{meas} \quad (1.40)$$

Input and output reproducibility matrices deviate from identity matrix. The choice of the damping factor defines the level of the solution approximation.

1.3.2 Regularization Using Regularization Matrix

This method due to Tikhonov [10], is an extension of the preceding method because it introduces a second regularization tool constituted by a matrix \mathbf{R} .

So the functional to minimize becomes:

$$J = (\mathbf{y}^{meas} - \mathbf{Ax})^T (\mathbf{y}^{meas} - \mathbf{Ax}) + \alpha[\mathbf{Rx}]^T [\mathbf{Rx}] \quad (1.41)$$

$$= \|\mathbf{y}^{meas} - \mathbf{Ax}\|^2 + \alpha\|\mathbf{R} \cdot \mathbf{x}\|^2 = \min \quad (1.42)$$

With this approach, a regularization condition (obtained with a matrix of data based on theoretical properties of the expected solution) is added to the preceding minimum. The result for the transformation matrix is the following:

$$\mathbf{A}^{-g} = (\mathbf{A}^T \mathbf{A} + \alpha \cdot \mathbf{R}^T \mathbf{R})^{-1} \cdot \mathbf{A}^T \quad (1.43)$$

Applications at case studies can clarify the criteria for the α parameter. For the matrix \mathbf{R} selection, a reference is given by the hole drilling method for residual stress measurement, one of the most ill-posed problems of stress analysis, Fig. 1.8, see discussion in [11, 12]. Figure 1.9 shows a synthesis of the inverse and pseudo-inverse matrices.

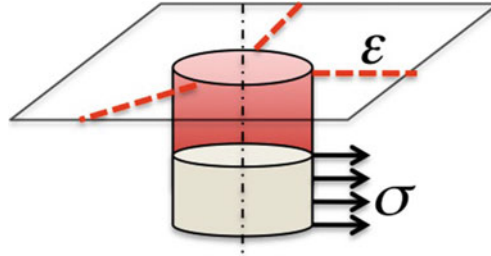


Fig. 1.8 The hole drilling method for the residual stress measurement, see ASTM E 837-08

$$\mathbf{A}^{-g} = \mathbf{A}^{-1} \quad \text{For even-posed problems}$$

$$\mathbf{A}^{-g} = \mathbf{A}^T (\mathbf{A} \cdot \mathbf{A}^T)^{-1} \quad \text{For under-posed problems}$$

$$\mathbf{A}^{-g} = (\mathbf{A}^T \mathbf{A})^{-1} \cdot \mathbf{A}^T \quad \text{For over-posed problems}$$

$$\mathbf{A}^{-g} = (\mathbf{A}^T \mathbf{A} + \alpha \cdot \mathbf{I})^{-1} \cdot \mathbf{A}^T \quad \text{Tikhonov-Phillips regularisation}$$

$$\mathbf{A}^{-g} = (\mathbf{A}^T \mathbf{A} + \alpha \cdot \mathbf{R}^T \mathbf{R})^{-1} \cdot \mathbf{A}^T \quad \text{Tikhonov-Miller regularisation}$$

Fig. 1.9 Inverse and pseudo-inverse matrices for the solution of inverse linear structural problems in matrix form [4]

1.3.3 Further Reading

The following sources can be useful to a reader seeking additional information about Inverse Problems [13–19]. For a bibliography of inverse problems in experimental solid mechanics see [20]. See also, for a general analysis on models in Engineering and Physics [21].

References

1. Friswell MI, Mottershead J (2001) Inverse method in structural health monitoring. *Key Engineering Materials Trans Tech Publ* 204–205:201–210
2. Di Barba P, Savini A, Wiak S (2008) *Field models in electricity and magnetism*. Springer, Dordrecht. ISBN 978-1-4020-6842-3
3. Doyle JF (2004) *Modern experimental stress analysis. Completing the solution of partially specified problems, vol 1*. Wiley, Hoboken. ISBN 0-470-86156-8
4. Laermann KH (2008) *Inverse problems in experimental structural analysis*. Shaker Verlag, Aachen
5. Liu GR, Han X (2003) *Computational inverse techniques in nondestructive evaluation*. CRC Press LLC, Boca Raton

6. Yaman F, Yakhno VG, Potthast R (2013) A survey on inverse problems for applied sciences. *Mathematical problems in engineering* (ID 976837)
7. Comandini M (2008) Progettazione con metodologie avanzate di organi di macchine sollecitati a fatica, corpi non saintvenantiani. Ph.D. thesis, University of Bologna, available as PDF file at <http://www.ingegneriindustriale.unibo.it/it/attivita-didattica/download-tesi-di-dottorato>
8. Gambini L (2007) Verifica a fatica di un braccio di escavatore con le teorie della tensione locale, della deformazione locale e relativo confronto con le principali normative. Master thesis, University of Bologna. [http://leonardo.gambini@libero.it](mailto:leonardo.gambini@libero.it)
9. Wolfram S (2012) *Mathematica manual*. Wolfram research, 8th edn. <http://www.wolfram.com/mathematica/>
10. Tikhonov AN, Arsenin WY (1977) *Solution of ill-posed problems*. Wiley, New York
11. ASTM (2013) E837-13a standard test method for determining residual stresses by the hole-drilling strain-gage method. ASTM international, West Conshohocken, PA, USA. www.astm.org
12. Casavola C, Pappalettere C, Tursi F (2010) Non-uniform residual stress fields on sintered materials. In: IGF (ed) *Proceedings 9th youth symposium on experimental solid mechanics, YSESM, Trieste*, pp 132–137. <http://www.gruppofrattura.it/sito/en/>
13. Bertero M, Boccacci P (1998) *Introduction to inverse problems in imaging*, vol 1, 1st edn. CRC Press (Application in imaging)
14. Groetsch CW (1999) *Inverse problems: activities for undergraduates*. Mathematical Association of America, Washington, DC. ISBN 0883857162 (pbk.)
15. Hansen PC (2010) *Discrete inverse problems: insight and algorithms*. Society for Industrial and Applied Mathematics, Philadelphia. <http://www.loc.gov/catdir/enhancements/fy1006/2009047057-t.html>
16. Idier J (2013) *Bayesian approach to inverse problems*, vol 1, 1st edn. Wiley, New York (Non-destructive evaluation)
17. Kirsch A (2011) *An introduction to the mathematical theory of inverse problems*, vol 1, 2nd edn. Springer Science & Business Media
18. Moura Neto FD, da Silva Neto AJ (2013) *An introduction to inverse problems with applications*, vol 1, 1st edn. Springer Science & Business Media
19. Tanaka M, Dulikravich GS (eds) (1998) *International symposium on inverse problems in engineering mechanics, Inverse problems in engineering mechanics (ISIP'98)*, vol 1, Nagano, Japan. ISBN: 978-0-08-043319-6
20. Pierron F, Avril S (2007) *Inverse problems in experimental solid mechanics*. *Experimental mechanics (Encyclopedia of Life Support Systems (EOLSS))*:39. <http://www.eolss.net/Eolss-sampleAllChapter.aspx>
21. Pfeiffer F, Bremer H (2015) *The art of modeling mechanical systems*. Course of CISM—AIMETA Advanced School, Udine. <http://www.cism.it/courses/C1507/>

Chapter 2

Introduction to the Application of Strain Gages

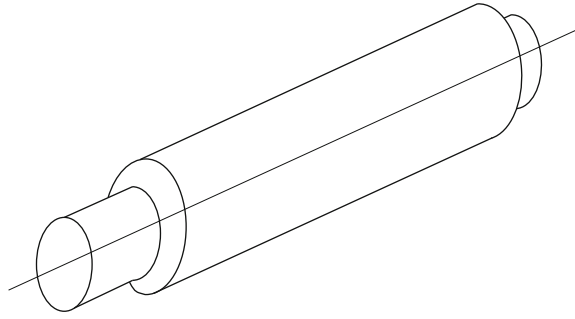
Abstract This chapter reviews the basic formulas of electrical resistance strain gages and related circuits, as well as some concepts that are not always emphasized in specialized books. Some sections of the chapter deal with the limits of the maximum and minimum values of the measurable strain, the choice of the resistance of the commercial gages series, the linearity and the drawbacks of the basic measuring circuits and some properties of a piezo-resistive gages. Classic formulas are given for principal strain and stress calculation for single strain gages and for rosettes, as well as for the application of the Wheatstone bridge circuit to biaxial strain state. A final issue concerns load cells. This topic is tackled from the point of view of design for high sensitivity and high stiffness which are especially required for fatigue tests. The calibration procedure is also described and applied to several cases, for uni- and multi-axial cells.

2.1 Properties of Strain Gages

The electrical resistance strain gage is the simplest and more widespread sensor developed for measuring static and dynamic strains at points of loaded structures and for the manufacturing of transducers and instruments of high accuracy and resolution. Several types of strain gages are offered on the market: the piezo-resistive type, the bonded metallic wire type, and foil resistance gages. The bonded resistance strain gage is by far the most widely used in experimental stress analysis as it will be shown; thus, the presentation is mostly devoted to this type. The manufacturers' catalogues and Technical Notes remain the most useful tools for the choice, e.g. [1–6], produced all over the world, in the USA, Germany, Japan and China. Nevertheless, some general guidelines can offer the basic description of the main characteristics for a first initiation among the great commercial variety of these sensors. For a complete presentation of the strain gages technique, see the books [7–12].

The principle of operation is based on a known property: an electrical conductor, subjected to strain along the axis, changes its electrical resistance. The strain can be tensile and compressive, (distinguished by a positive or negative sign) representing expansion or contraction respectively.

Fig. 2.1 Changes of length and diameter of an electrical conductor when it is stretched or compressed



When the conductor, electrically isolated by a thin backing called carrier, is bonded in a certain direction to a restricted area of the test element, from the measurement of the change in resistance it is possible to discover the strain in that direction.¹ This distance is the gage length and is an important parameter for a comparison between various strain measurement techniques.

The measurement of a change in resistance between two points of an electrical conductor close to one another is much more accurate than the direct measurement of the variation of the distance between the same points obtained by mechanical devices; furthermore, being an electrical signal, it offers enormous advantages from the point of view of sensitivity, resolution, conditioning, storing, comparison and remote control capabilities.

The minimal significant variation of the distance (related to the initial value) is of the order of magnitude of $10^{-6} \frac{\text{m}}{\text{m}}$ or $1 \frac{\mu\text{m}}{\text{m}}$. This unit, common in practical use, is called *micro-strain* ($\mu\epsilon$) and is utilized for expressing deformation values.²

2.1.1 Relationship Between Strain and Resistance Change

An electrical conductor (not necessarily of circular section), subjected to deformation along its axis, undergoes geometric variations (length and cross-section), Fig. 2.1 and a physical variation due to the change of its resistive property.

For a conductor of cross-section S and length l the electrical resistance is:

$$R = \frac{\rho l}{S} \quad (2.1)$$

Taking the logarithms of the two members of Eq. 2.1 and the derivatives of them with respect to an auxiliary variable t , we have:

¹ Strain is defined as the displacement between two points some distance apart, related to the initial distance.

² For a typical average strain value in a steel structure of $10^{-4} \text{ m/m} = 10^{-4} \text{ mm/mm} = 10^{-4} \text{ in/in}$ the equivalent values in this unit are of $10^{-2} \%$ or $1000 \mu\epsilon$.

$$\log R = \log \rho + \log l - \log S = \log \rho + \log l - 2 \log \Phi - \log \frac{\pi}{4} \quad (2.2)$$

$$\frac{1}{R} \frac{dR}{dt} = \frac{1}{\rho} \frac{d\rho}{dt} + \frac{1}{l} \frac{dl}{dt} - \frac{2}{\Phi} \frac{d\Phi}{dt}$$

Eliminating the common auxiliary increment dt , the following relationship remains:

$$\frac{dR}{R} = \frac{d\rho}{\rho} + \frac{dl}{l} - 2 \frac{d\Phi}{\Phi}$$

The first term $d\rho/\rho$ is due to the change in resistivity and can be interpreted as a small piezo-resistive effect, while the terms $\varepsilon_l = dl/l$ and $\varepsilon_t = d\Phi/\Phi = -\nu\varepsilon_l = -\nu dl/l$, due to geometrical variations of the conductor in longitudinal and in lateral directions, are the dominant effect for metallic materials. Then:

$$\frac{dR}{R} = \frac{d\rho}{\rho} + \frac{dl}{l}(1 + 2\nu) \quad (2.3)$$

Collecting dl/l :

$$\frac{dR}{R} = \left[1 + 2\nu + \frac{d\rho/\rho}{dl/l} \right] \frac{dl}{l}$$

and substituting infinitesimal variations with finite increments, we have:

$$\frac{\Delta R}{R} = S\varepsilon \quad (2.4)$$

where

$$S = \left[1 + 2\nu + \frac{d\rho/\rho}{dl/l} \right]$$

It can be concluded that if S is constant, the relative variation (called also fractional change) in resistance is proportional to the fractional change in length.

S represents the conductor sensitivity. Since the geometric contribution alone is, for every metallic alloy, between $(1 + 2\nu) = 1.5 \div 1.7$ and becomes 2 in the plastic field, while the measured value of S for metals for elastic (reversible) behavior is between 2 and 4, it is possible to conclude that $d\rho/\rho$ due to a piezo-resistance effect of the wire material is small but not negligible, and for special materials can be even a big term.

Table 2.1 Example of commercial strain gages materials (Vishay precision group M.M. [6])

Material	Composition	Sensitivity	Use
Constantan	45Ni 55Cu	2.1	General
Karma	73 Ni; 20Cr; Fe and Al	2.1	General
Isolastic	36.5 Ni; 8Cr; 0.5 Mo; 55 Fe	3.6	Dynamic
Nichrome	80 Ni; 20 Cr	2.1	$T > 250\text{ }^{\circ}\text{C}$
Platinum-Tungsten	92 Pt; 8 W	4.0	$T > 250\text{ }^{\circ}\text{C}$

Isoelastic, Constantan, Nichrome, Karma, etc. are trade names

2.1.2 Materials for Metal Strain Gages

Alloys used for strain gages manufacturing must exhibit a linear relationship between the relative change of resistance and the strain, as well as a low temperature coefficient of the electrical resistance. Table 2.1 shows the commercial names, chemical composition, sensitivity and field of application of five of the most common alloy materials for strain gages produced by a specialized Company [6] that satisfy the previous conditions and other requirements relating to temperature stability, to fatigue strength and to other special applications in hostile environment.³ Other manufacturers [5] offer similar types of Strain gages that cover typical applications:

- Universal strain gages for stress analysis and simple transducers with a grid in Constantan and carrier in Polyimide.
- Special strain gages for measurements at extreme temperatures from -269 up to $250\text{ }^{\circ}\text{C}$ with a grid in Cr–Ni alloy and carrier in Polyimide.

Limiting the analysis to some of them, *Constantan* is the most widely used and the least expensive gage wire material. This material is suitable for static or quasi-static strain applications, for large deformations, even if it is not suitable for high temperatures. This material is particularly attractive, because the geometrical contribution of the conductor to the relative variation of the resistance (i.e. the sensitivity) is 2 and it maintains its value for the entire deformation interval, Fig. 2.2a. It is stable with respect to thermal variation so that the ratio

$$\frac{d\rho/\rho}{\varepsilon}$$

is relatively constant at different values of ε , where $\varepsilon = dl/l$. Sensors can be built with a linear response up to great deformations; only at very large deformations such as $150,000\text{ }\mu\text{m/m} = 15\%$, is a non-linearity characteristic found and the

³ These are commercial names, but, due to the reputation of this Company, can constitute a reference also for others that use similar alloys with different commercial names.

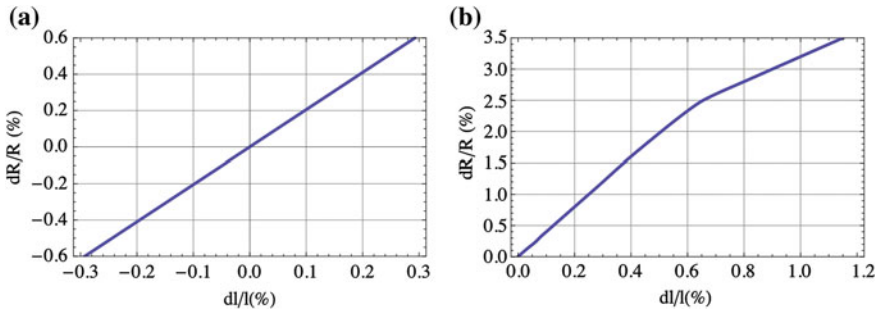


Fig. 2.2 Schematic representation of the resistance versus strain characteristics of tensioned wires of Constantan (a) and a Pt alloy (b)

relationship becomes non linear (parabolic). More sensitive sensors can be built with other materials but they are much more critical with respect to the linearity.

With a proper manufacturing procedure of the alloy and of the selected melt, the gage wire manifests a very low thermally induced strain (apparent strain) over a wide range of temperatures. The same characteristics are typical of other alloys such as *Karma* alloy, so that this material, together with Constantan, is considered to have self-temperature-compensation.

In the field of *FeNiCr* alloys, the *Isoelastic alloy* has a higher sensitivity (3.6 versus 2.1 in Constantan) but, unlike Constantan, does not have the self-temperature-compensation property and is too sensitive to changes in temperature, so that it is not suitable for measurements with temperature variation over time.

Since most strain gages built with this alloy have a resistance of 350 Ω (compared to 120 Ω of Constantan strain gages), the signal to noise ratio is improved.

However, the real advantage of this alloy is concerned with the better fatigue properties with respect to other strain gage materials, even if the sensitivity remains not constant with respect to deformation or to temperature and decreases from 3.5 to 2.5 when the strain exceeds a threshold of 8,000 $\mu\epsilon$. Concluding, this alloy is suitable for standard dynamic strain measurement in vibration and impact applications.

For *Chrome-Nickel and Platinum-Tungsten alloys* or similar alloys of other manufacturers,⁴ S is not constant with respect to the elongation and the temperature, Fig. 2.2b. Nevertheless, they are resistant to high temperatures, and are thus used for $T > 250^\circ\text{C}$ or for other special purposes.

⁴ For HBM, these gages have the specifications of Series C.

2.1.3 Resistance Values

The typical values of the strain gages on the market are 120, 350, 600, 1000 Ω .⁵ Let us examine the reason for this manufacturer choice. The elastic strains in a structural mechanics problem lie in the interval of $10^{-4} \div 1\%$. This means that, e.g. for a steel bar in pure tension (with the elastic modulus of about 200,000 MPa) this interval corresponds to a stress range from 0.2 to 2,000 MPa.

With a resistance value of 120 Ω , and S equal to 2, the minimum value of the sensitivity, the resistance varies in the range:

$$2.4 \times 10^{-4} \Omega \leq \Delta R \leq 2.4 \Omega$$

The instruments available today have a measuring range that fits the previous one, with a *sensitivity* equal to:

$$\frac{1 \text{ division}}{10^{-4} \Omega}$$

The minimum value thus corresponds to 2.4 divisions, an acceptable resolution for measurements of very small deformations. Modern instruments have optimum *linearity* and adequate *class of accuracy*.⁶

The value of about 120 Ω is, generally, the lower commercial limit for a strain gage resistance. Through the wire technology, this resistance requires a minimum value of the conductor length of about 120 mm.⁷ Also with modern techniques of photoengraving, that make it possible to produce very thin gages on a polyimide resin carrier, approximately 13 μm thick, the conductor length is not negligible. Moreover, if larger values of resistance (e.g. 350 Ω) are preferable for the increment of the measurable resistance variation with a consequent reduction of the *noise to signal* ratio, such resistances are limited by the maximum acceptable length of the conductor. A maximum resistance value of 1,000 Ω can be regarded as the upper limit in extreme applications.

⁵ Other less common values of the gages [13] with a grid in Constantan, can reach the value of 60 Ω .

⁶ Often the terms resolution, precision, and accuracy are used interchangeably, but they actually indicate very different entities: resolution is the fineness to which an instrument can be read and precision is the fineness to which an instrument can be read with a good repeatability and reliability.

The class of accuracy is the percentage of the inherent error of the measuring device with respect to full scale indication. E.g., if the class of accuracy is 2, this means an error of 2 in a full scale 100 reading [14, 15].

⁷ For an electrical conductor of $\Phi = 25 \mu\text{m} = 0.025 \text{ mm}$ in diameter and $\rho = 0.49 \Omega \frac{\text{mm}^2}{\text{m}} = 0.49 \times 10^{-3} \Omega \text{ mm}$ for Constantan:

$$l_{min} = 120 \cdot \frac{4.9 \times 10^{-4}}{4.9 \times 10^{-4}} = 120 \text{ mm}$$

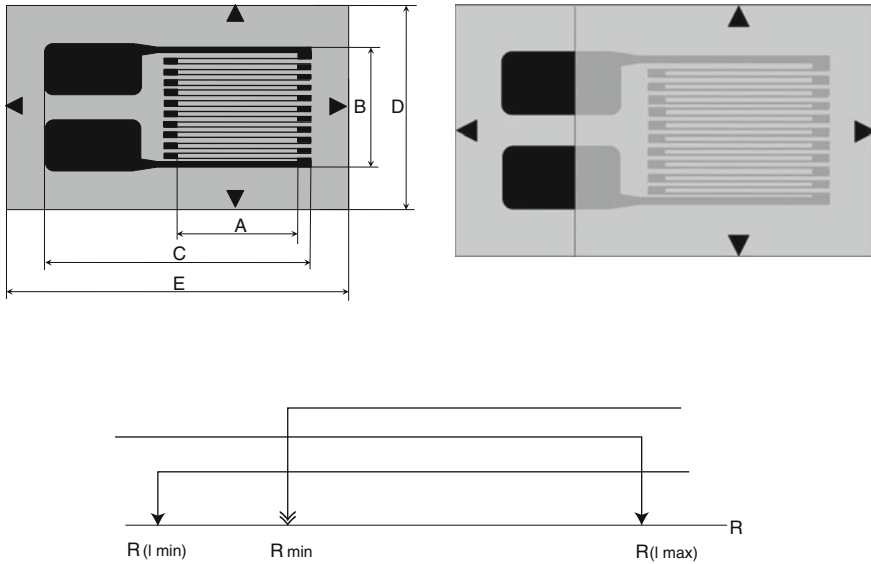


Fig. 2.3 Sketches of a Normal (on the *left*) and Encapsulated (on the *right*) thin Film Strain gage and its resistance range. A = gage length, B = grid width, C = overall pattern, D = matrix width, E = matrix length

The reason for the adopted solution for the gages, in which the conductor is wound (today engraved) in a grid with a serpentine shape, is now clear. All segments placed along a direction cooperate in the overall resistance, while the *gage length* i.e. the measuring base, coincides with the length of a segment, and is the distance over which the measured strain is averaged. Thus in areas with high strain gradients, the grid length must be very small, which greatly affects the conductor length, Fig. 2.3, The thin isolated support must, when bonded, guarantee the perfect transmission of the deformation from the measuring point to the grid.

Manufacturing the resistance strain gage by a photoengraving and etching process allows grids of thickness 3.8 to 5 μm . For resistance values of 120 Ω grids with very small length from 0.2mm are thus available. There is another constraint on the minimum value of the resistance due to the capacity to dissipate power in the form of thermal energy (heat). This limit can be higher than the commercial value of the strain gage, but a universal indication for this limit is not possible, because factors other than gage type are involved.

All resistors have a steady-state power rating, which is the amount of Joule power that they can dissipate for an indefinite time, without causing damage by overheating. It depends on what other materials the grid is in contact with (conductive cooling), the grid area,⁸ the carrier type and the adhesive properties, the thermal capacity of body, the airflow around the conductor, etc.

⁸ The grid area is the active gage length multiplied by the grid width.

Table 2.2 Maximum acceptable grid power-density, from [9]

Tested body material	P_D (mW/mm ²)	Mean value
Thick Al or Cu	8 ÷ 16	12
Thick steel	3 ÷ 8	5
Thin steel	1.5 ÷ 3	2
Filled plastic or ceramic	0.3 ÷ 0.8	0.5
Unfilled plastic	0.3 ÷ 0.8	0.05

The not dissipated power, that is the second portion of the thermal power, is responsible for a temperature increment so that both grid and substrate operate at temperatures higher than the room temperature. When the temperature rise is excessive, gage performance is affected by a variation of the electrical resistance.

The admitted temperature must be limited by this constraint, as well as by creep effects. Experience [9] suggests a maximum value for the total Joule thermal power P_D (mW/mm²) per unit area of the grid, a function of the thermal capacity of the body and of its thermal conductivity, Table 2.2.

The thermal power varies as the square of the applied voltage: $P_W = R \cdot i^2 = e^2/R$, where e is the voltage at the strain gage terminals. This power must not be bigger than the maximum power dissipated by the conduction effect which is a function of the heat-sink capacity of the body.

This limit is the product of the maximum dissipated power per unit area and the area A of the grid:

$$P_W \leq P_D \cdot A$$

The total power is a function of e , thus a second minimum condition for the strain gage resistance is:

$$R \geq \frac{e^2}{P_D \cdot A} \quad (2.5)$$

The e values, functions of the supply voltage e_0 of the measuring circuit, can be found by a trial and error method, by gradually increasing the bridge excitation⁹ under zero-load conditions until a zero instability is observed. The excitation should then be reduced until the zero reading becomes stable again, without a significant offset.

Following the suggestion of a manufacturer [6]: “Conducting this test at the maximum operating temperature instead of room temperature will increase the likelihood that the maximum safe bridge voltage has been established.”

Commercial strain measurement instrumentation utilizes constant excitation voltage e_0 of 1 to 10 V, (with mean values of 1 ÷ 2 V), thus the voltage across the active arm e (one-half the bridge voltage for a bridge with equal resistances) varies from

⁹ As it will be shown, the Wheatstone bridge is the most used measuring circuit for the strain gages. In a normal set-up the bridge has four equal resistances on its arms.

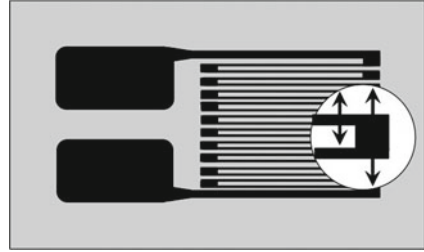
Table 2.3 Minimum gage resistance values for the admissible power densities

e_0 (V) Bridge	e (V) Gage	Body material	P_D ($\frac{mW}{mm^2}$)	$R(\Omega)$ Grid area = 3×1.4	$R(\Omega)$ Grid area = 6×2.8
1	0.5	Thick Al or Cu	$8 \div 16$	120	120
		Thick steel	$3 \div 8$	120	120
		Thin steel	$1.5 \div 3$	120	120
		Filled plastic or ceramic	$0.3 \div 0.8$	120	120
		Unfilled plastic	$0.03 \div 0.08$	1000	350
2	1	Thick Al or Cu	$8 \div 16$	120	120
		Thick steel	$3 \div 8$	120	120
		Thin steel	$1.5 \div 3$	120	120
		Filled plastic or ceramic	$0.3 \div 0.8$	600	120
		Unfilled plastic	$0.03 \div 0.08$	>1000	>1000
10	5	Thick Al or Cu	$8 \div 16$	600	120
		Thick steel	$3 \div 8$	>1000	350
		Thin steel	$1.5 \div 3$	> 1000	600
		Filled plastic or ceramic	$0.3 \div 0.8$	>1000	>1000
		Unfilled plastic	$0.03 \div 0.08$	>1000	>1000

0.5 to 5 V. According to Eq. 2.5 the minimum values of the strain gage resistance can be calculated as a function of the bridge voltage e_0 , for different grid areas and heat-sink properties of the gage type, Table 2.3. From the Table values the following remarks can be made on the excitation voltage:

- For all the cases for which the resistance exceeds the commercial values, the excitation voltage must be reduced, selecting a gage with a larger area, or reducing the bridge voltage with a potentiometric circuit, Fig. 2.8 or a series resistor.
- Creep in the gage backing and adhesive will occur also at room temperature, when the temperatures of grid and substrate are raised by self-heating effects.
- Output due to temperature will be altered when grid and substrate temperatures are significantly different.
- Gages for normal stress analysis can be excited at a higher level than under transducer conditions, where the utmost in stability, accuracy, and repeatability is very high.

Fig. 2.4 Transversal strain on longitudinal grid elements and on the inversion portions of the conductor



- Gages for dynamic strain measurement can be excited by higher voltage with a considerable advantage on measurement accuracy.

A great number of strain gage types are available on the market for the most widely differing strain measurement applications, from experimental stress analysis applications to transducer development, for a variety of resistance values, lengths, patterns, alloy materials to stainless steel, carbon steel, and aluminum (matched thermal expansion coefficients ($\alpha = 5, 12, 16, 23, 27 \text{ ppm}/^\circ\text{K}$), substrates, adhesives and bonding, and solder configurations. Several companies also provide installation accessories and instruments for test and measurement applications. New inventions related to manufacturing methods are continually being proposed as one based on an ion beam technology.¹⁰

2.1.4 Transverse Sensitivity and Strain Gage Factor

In the serpentine arrangement of the strain gage bonded on a calibration specimen in a longitudinal direction, the longitudinal grid elements and the inversion portions of the conductor are also stretched in a transversal direction, Fig. 2.4 and the total resistance variation of the gage is also a function of these minor effects. Remembering that:

$$R_l + R_t = R$$

the total resistance variation for a strain gage is:

$$\Delta R = S \cdot \varepsilon_l R_l + S \cdot \varepsilon_t R_t$$

thus:

$$\frac{\Delta R}{R} = S \cdot \varepsilon_l R_l \cdot \frac{1}{R} + S \cdot \varepsilon_t R_t \cdot \frac{1}{R}$$

¹⁰ The method is described in (CN 101614522 B patent on Manufacturing method of resistance strain gage based on ion beam technology (02–Mar–2011) Applicants: NO 44 Inst. of No 4 Academy of China Aerospace Science and technology Corp. 866.473.6826, <http://ip.com/>).

If:

$S_l = S \cdot R_l/R$ is the longitudinal sensitivity

$S_t = S \cdot R_t/R$ is the transversal sensitivity

$$\frac{\Delta R}{R} = S_l \cdot \varepsilon_l + S_t \cdot \varepsilon_t = S_l (\varepsilon_l + K_t \varepsilon_t) \tag{2.6}$$

with $K_t = \frac{S_t}{S_l}$.

If the strain gage is bonded in a longitudinal direction on a specimen with a known bending moment Fig. 2.5, holds:

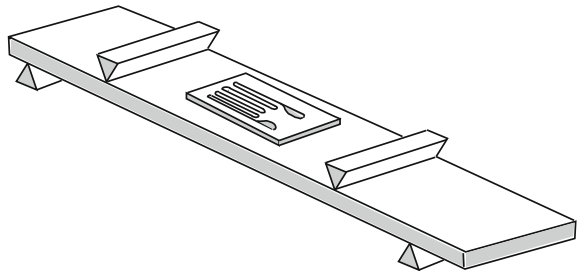
$$\frac{\Delta R}{R} = S_l (\varepsilon_l + K_t \varepsilon_t) = S_l (\varepsilon_l - K_t \nu_0 \varepsilon_l) = S_l \varepsilon_l (1 - \nu_0 K_t) = K \cdot \varepsilon$$

$K = S_l (1 - \nu_0 K_t)$ is the gage factor, i.e. the proportionality factor between the fractional resistance change and the longitudinal strain. The factor also takes into account the backing and glue effect that partially shield the strain transition from the body to the grid. As the constant was defined, the calibration of the strain gage factor K depends on the ratio between longitudinal and transverse strains of the calibration bar. A systematic error arises in all the cases of strain gage applications in a biaxial stress field, with a ratio other than that used in the calibration specimen [16]. Even if errors in strain indication due to transverse sensitivity are generally small since the transverse sensitivity is small, (in common gages it varies from -9.2 to $+1.8$ %), these errors can become intolerable in biaxial strain fields with large differences between principal strains [17]. It can be taken into account and corrected in a simple way for two gages bonded in whatever perpendicular directions 1 and 2 of a biaxial state of stress, where the fractional gage resistances in both directions can be written as:

$$\frac{\Delta R_1}{R_1} = S_l \cdot \varepsilon_1 + S_t \cdot \varepsilon_2 = K \hat{\varepsilon}_1$$

$$\frac{\Delta R_2}{R_2} = S_l \cdot \varepsilon_2 + S_t \cdot \varepsilon_1 = K \hat{\varepsilon}_2$$

Fig. 2.5 Device for gage calibration



where: $\hat{\varepsilon}_1$ and $\hat{\varepsilon}_2$ are the indicated not corrected values with the gage factor K . Substituting its expression leads to the equations:

$$\begin{cases} S_l (\varepsilon_1 + K_t \varepsilon_2) = S_l (1 - \nu_0 K_t) \hat{\varepsilon}_1 \\ S_l (\varepsilon_2 + K_t \varepsilon_1) = S_l (1 - \nu_0 K_t) \hat{\varepsilon}_2 \end{cases}$$

from which the *true* strains in both directions 1 and 2 are derived:

$$\begin{cases} \varepsilon_1 = \frac{1 - \nu_0 K_t}{1 - K_t^2} (\hat{\varepsilon}_1 - K_t \hat{\varepsilon}_2) \\ \varepsilon_2 = \frac{1 - \nu_0 K_t}{1 - K_t^2} (\hat{\varepsilon}_2 - K_t \hat{\varepsilon}_1) \end{cases} \quad (2.7)$$

2.1.5 Influence of a Temperature Variation

It is known that the resistance of an electrical conductor R is strongly affected by the temperature; so the strain gage grid, the lead wiring, as well as the backing and the body material to which the gage is bonded, are strongly influenced by changes in temperature, these variations constituting the most common causes of error of the measurements.

Only a temperature variation of the bonded gages *during the test* influences the results. A resistance variation before the bridge balance is ineffective and compensated and does not influence the measurements. Vice-versa, during-the-test temperature changes are causes of relevant errors. Let us consider the thermal phenomenon that affects all the variables, inducing:

- A resistivity variation of the grid material, expressed in a linearized form for small thermal variations, as:

$$\rho_t = \rho_0 (1 + \gamma \Delta T)$$

then

$$\frac{\Delta \rho}{\rho_0} = \gamma \Delta T$$

This fractional variation takes into account the modification of the resistance of a unitarian dimension of the conductor material, due to two reasons:

- the geometrical variation of the unitarian volume (length and cross-section of the conductor) and
- the modification of the physical material property.

As a consequence, the fractional resistance variation is the same as the fractional resistivity variation. Indeed:

$$\frac{\Delta R}{R} = \Delta \rho \frac{l}{S} \frac{1}{R} = \frac{\Delta \rho}{\rho} = \gamma \Delta T$$

This term is the first contribution of an *apparent* strain:

$$\varepsilon_{app} = \frac{\gamma \Delta T}{K} \quad (2.8)$$

(e.g. for Constantan $\gamma = 40 \text{ ppm}/^\circ\text{K}$)

- In addition, a sensitive grid length variation must be considered, equal to:

$$l_t = l_0(1 + \alpha \Delta T)$$

(for the Constantan $\alpha = 15 \text{ ppm}/^\circ\text{K}$)

And the fractional contribution to the apparent strain is:

$$\varepsilon_{app} = \frac{l_t - l_0}{l_0} = \alpha \Delta T$$

- Moreover, a length variation of the portion of the body on which the gage is bonded occurs for the linear dilatability coefficient of the tested objects:

$$l_t = l_0(1 + \beta \Delta T)$$

β is, e.g. equal to $12 \text{ ppm}/^\circ\text{K}$ for Carbon steel $16 \text{ ppm}/^\circ\text{K}$ for inox austenitic steel $24 \text{ ppm}/^\circ\text{K}$ per aluminum alloys.

The fractional contribution to the apparent strain is:

$$\varepsilon_{app} = \beta \Delta T \quad (2.9)$$

- The expression of the total apparent strain is thus:

$$\varepsilon_{app} = \left[\frac{\gamma}{K} + (\beta - \alpha) \right] \Delta T \quad (2.10)$$

The minus sign is justified by the deformation sign conventions. If $\beta > \alpha$ this strain shows an apparent tension and a compression in the contrary case.

- The gage factor is finally influenced by a temperature change, according to a linear law, as follows:

$$K_T = K_0(1 + \delta \Delta T) \quad (2.11)$$

with a value for the most common gages equal to $\delta = 95 \text{ ppm}/^\circ\text{K}$. The real value is given by the manufacturer in the strain gage packaging, together with the gage factor and a complete diagram of the apparent strain for the actual base materials.

The previous relationship shows that the apparent strain depends not only on the nature of the strain gage, but also on the material to which the gage is bonded. For this reason, thermal output data are meaningful only when referred to a particular type of strain gage, bonded to a specified material.

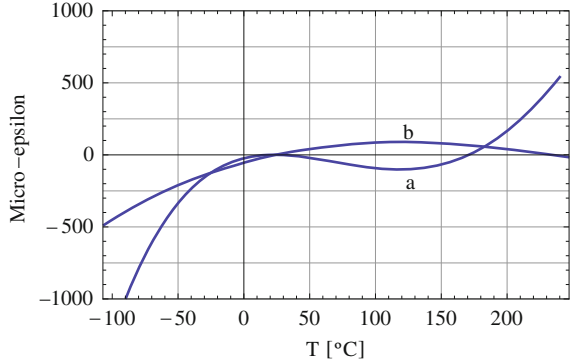
- In order to reduce the apparent strain of Eq. 2.10 it is thus necessary to select the coefficients α , β and γ in order to reduce it to a minimum value or better to cancel it completely. Manufacturers suggest, the coefficient γ of the grid material being known, the proper choice of α for matching the body material coefficient β , in order to have a global compensation on the apparent strain. The gages of this kind are called self-compensating gages.
- In the case of dynamic or cyclic strain measurements, if the scope of the measurements is the strain variation at a frequency higher than the frequency of temperature change, the thermal output (apparent strains) on the mean signal value is theoretically of no consequence on the measurement results, and it can be neglected.
- Nevertheless, the errors due to thermal output can not be compensated on the entire temperature range but only on a limited interval close to the room temperature, but they become extremely large as temperatures deviate from the reference temperature during tests.

2.1.6 Compensation for Thermal Output

The way to reduce apparent strains is the use of self-temperature compensation gages and/or utilizing proper circuit arrangements. In order to keep apparent strain as small as possible (at least in the normal range of working temperatures) in case of temperature changes, the metallurgical properties of grid materials are modified by adding alloying elements and by working them with proper thermo-mechanical treatment.

Strain gage manufacturers [18] provide, e.g. for each lot of the most common alloy gages, diagrams like that in Fig. 2.6 for self-temperature compensated Constantan (A-alloy) and modified Karma (Vishay K-alloy) strain gages. These diagrams show the effect of the self-compensation when the gages are bonded to material having the thermal expansion coefficient for which the gage is intended. A regression-fitted (least-squares) polynomial equation is given to calculate the apparent strain of Constantan alloy gages. This equation is useful for temperatures outside the field of a total compensation.

Fig. 2.6 Thermal output variation with temperature for self-temperature compensated, (a) Constantan and (b) (Vishay K-alloy) strain gages



$$\varepsilon_T = A_0 + A_1T + A_2T^2 + A_3T^3 + A_4T^4 \tag{2.12}$$

where T is the grid temperature during the test.

For Constantan alloy, the following average coefficients can be used for representing the curve with T in Fahrenheit degrees:

$$\varepsilon = -8.54 \times 10 + 2.65 \cdot T - 2.47 \times 10^{-2} \cdot T^2 + 6.50 \times 10^{-5} \cdot T^3 - 3.84 \times 10^{-8} \cdot T^4$$

or the following for T in Centigrade degrees, Fig. 2.6:

$$\varepsilon = -23.803 + 2.2749 \cdot C - 0.060575 \cdot C^2 + 0.00035041 \cdot C^3 - 4.0311 \times 10^{-7} \cdot C^4$$

Zero of the function is for T equal to the ambient temperature.

Those equations are obtained for a gage factor equal 2.0, so that apparent strains calculated by the previous equations must be adjusted to a different gage factor.

The correction on the gage factor that also takes into account the thermal effect computed in Eq. 2.11 due to backing and glue materials is evaluated by the following Eq. 2.13:

$$\varepsilon = \hat{\varepsilon} \cdot \frac{K_0}{K(T)} \tag{2.13}$$

where:

ε is the corrected strain,

K_0 is the calibration factor equal 2.0 for Constantan,

$K(T)$ the actual value at the test temperature,

$\hat{\varepsilon}$ the indicated (not corrected) strain corresponding to $K = 2$.

The apparent strain must be algebraically subtracted from the indicated (not corrected) strain on the display of the instrument.

Conversely, it is clear that, if the gage is used within the range of the compensation, it is not necessary to compensate its thermal output.

2.2 Strain Gage Rosettes

According to the elastic theory, the strain value in a point ε_θ in the direction θ can be expressed as a function of the normal and shear strain in two mutual perpendicular axes x, y .

- the strain $\varepsilon_x = dx/x$ gives this contribution along θ direction:

$$\varepsilon_\theta = \frac{da}{a} = \frac{dx \cos \theta}{x / \cos \theta} = \varepsilon_x \cos^2 \theta \quad (2.14)$$

- the strain $\varepsilon_y = dy/y$ gives this contribution along θ direction:

$$\varepsilon_\theta = \frac{da}{a} = \frac{dy \sin \theta}{y / \sin \theta} = \varepsilon_y \sin^2 \theta$$

- the shear strain γ_{xy} , gives this contribution along θ direction:

$$\varepsilon_\theta = \frac{da}{a} = \frac{\gamma_{xy} a \sin \theta \cos \theta}{a} = \gamma_{xy} \sin \theta \cos \theta = \frac{\gamma_{xy}}{2} \sin 2\theta$$

Thus the total strain along θ direction is:

$$\varepsilon_\theta = \varepsilon_x \cos^2 \theta + \varepsilon_y \sin^2 \theta + \frac{\gamma_{xy}}{2} \sin 2\theta$$

or:

$$\varepsilon_\theta = \frac{\varepsilon_x + \varepsilon_y}{2} + \frac{\varepsilon_x - \varepsilon_y}{2} \cos 2\theta + \frac{\gamma_{xy}}{2} \sin 2\theta \quad (2.15)$$

Conversely, in order to derive the normal and shear strains from the strain ε_θ at a point $\varepsilon_x \varepsilon_y \gamma_{xy}$, referring to a coordinate system of cartesian axes, three measurements of ε_θ in three different θ directions are needed:

$$\begin{cases} \varepsilon_a = \frac{\varepsilon_x + \varepsilon_y}{2} + \frac{\varepsilon_x - \varepsilon_y}{2} \cos 2\theta_a + \frac{\gamma_{xy}}{2} \sin 2\theta_a \\ \varepsilon_b = \frac{\varepsilon_x + \varepsilon_y}{2} + \frac{\varepsilon_x - \varepsilon_y}{2} \cos 2\theta_b + \frac{\gamma_{xy}}{2} \sin 2\theta_b \\ \varepsilon_c = \frac{\varepsilon_x + \varepsilon_y}{2} + \frac{\varepsilon_x - \varepsilon_y}{2} \cos 2\theta_c + \frac{\gamma_{xy}}{2} \sin 2\theta_c \end{cases} \quad (2.16)$$

The idea of *rosette gages* derives from these Equations.

Having a single sensor formed by three gages assembled in optimal directions on the same backing facilitates the bonding and the electrical connection of two/three gages, Fig. 2.7. Stacked Rosette gages allow a more accurate measurement at the same point but are critical from the point of view of the thermal power and of the

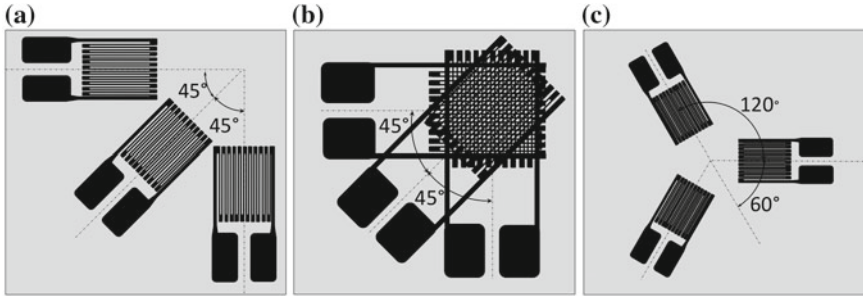


Fig. 2.7 Schematic representation of three-gage Rosettes with different layout of the strain gages at 45 and 60/120° with separated or stacked grids

local reinforcement effects.¹¹ It follows that a single gage with one input power will have a temperature rise at the surface equal to double the mean temperature of the grid, (taking the temperature of the specimen as equal to room temperature). Then in a stacked three-gage rosette similarly excited, the top grid will receive six times the temperature rise of a similar single gage.

In this case of the three gages, the power applied to each gage must be reduced to 1/6, to keep the temperature rise of the top grid equal to that of a single gage. Since the power is proportional to the square of the applied voltage ($P_W = e^2/R$), the excitation voltage must be reduced to $1/\sqrt{6} = 1/2.5$.

If the strain gradient at the measuring point is not too high and the deformation is essentially constant in the area around the point, rosettes with not overlapped gages are the appropriate solution.

2.2.1 Three-Gage Rectangular (0° 45° 90°) Rosettes

In this case the θ values are:

$$\begin{cases} \theta_a = 0^\circ \\ \theta_b = 45^\circ \\ \theta_c = 90^\circ \end{cases} \quad (2.17)$$

¹¹ For Fourier’s law $q_x = -kAdT/dx$ holds, where: q_x is the amount of heat transferred per unit time that is constant for each gage.

k is the thermal conductivity.

A is the exchange area.

From $dT/dx = const.$ a linear law of temperature variation along the gage thickness is derived.

The Eq. 2.16 give:

$$\begin{cases} \varepsilon_a = \frac{\varepsilon_x + \varepsilon_y}{2} + \frac{\varepsilon_x - \varepsilon_y}{2} = \varepsilon_x \\ \varepsilon_b = \frac{\varepsilon_x + \varepsilon_y}{2} + \frac{\gamma_{xy}}{2} \\ \varepsilon_c = \frac{\varepsilon_x + \varepsilon_y}{2} - \frac{\varepsilon_x - \varepsilon_y}{2} = \varepsilon_y \end{cases}$$

then:

$$\begin{cases} \varepsilon_x = \varepsilon_a \\ \gamma_{xy} = 2\varepsilon_b - \varepsilon_a - \varepsilon_c \\ \varepsilon_y = \varepsilon_c \end{cases}$$

and for the principal strains:

$$\begin{cases} \varepsilon_1, \varepsilon_2 = \frac{\varepsilon_x + \varepsilon_y}{2} \pm \sqrt{\left(\frac{\varepsilon_x - \varepsilon_y}{2}\right)^2 + \left(\frac{\gamma_{xy}}{2}\right)^2} \\ \tan 2\phi = \frac{\gamma_{xy}}{\varepsilon_x - \varepsilon_y} \end{cases}$$

or:

$$\begin{cases} \varepsilon_1, \varepsilon_2 = \frac{\varepsilon_a + \varepsilon_c}{2} \pm \sqrt{\left(\frac{\varepsilon_a - \varepsilon_c}{2}\right)^2 + \left(\frac{2\varepsilon_b - \varepsilon_a - \varepsilon_c}{2}\right)^2} \\ \tan 2\phi = \frac{2\varepsilon_b - \varepsilon_a - \varepsilon_c}{\varepsilon_a - \varepsilon_c} \end{cases} \quad (2.18)$$

ϕ is the angle between the principal direction 1 and the x axis.

The determination of the actual value of ϕ from the previous equation needs the signs analysis of the numerator z and denominator n of the Equation:

$$\phi = \frac{1}{2} \arctan \frac{z}{n} \quad (2.19)$$

Table 2.4 shows the actual angle values for different signs of the numerator and denominator. The corresponding state of stress is determined by the constitutive law, e.g. the *Hooke's* law, valid for the elastic-isotropic case:

Table 2.4 Determination of the actual directions of the principal strains

z	≥ 0	> 0	≤ 0	< 0
n	> 0	≤ 0	< 0	≥ 0
ϕ	$\phi = \phi^*$ $0^\circ \leq \phi < 45^\circ$	$\phi = \frac{\pi}{2} - \phi^*$ $45^\circ \leq \phi < 90^\circ$	$\phi = \frac{\pi}{2} + \phi^*$ $90^\circ \leq \phi < 135^\circ$	$\phi = \pi - \phi^*$ $135^\circ \leq \phi < 180^\circ$

$$\begin{cases} \sigma_1 = \frac{E}{1-\nu^2}(\varepsilon_1 + \nu\varepsilon_2) \\ \sigma_2 = \frac{E}{1-\nu^2}(\varepsilon_2 + \nu\varepsilon_1) \end{cases} \quad (2.20)$$

Substituting Eq. 2.18 gives:

$$\begin{cases} \sigma_1, \sigma_2 = \frac{E}{2} \left[\frac{\varepsilon_a + \varepsilon_c}{1-\nu} \pm \frac{1}{1+\nu} \sqrt{(\varepsilon_a - \varepsilon_c)^2 + (2\varepsilon_b - \varepsilon_a - \varepsilon_c)^2} \right] \\ \phi = \frac{1}{2} \arctan \frac{2\varepsilon_b - \varepsilon_a - \varepsilon_c}{\varepsilon_a - \varepsilon_c} \end{cases} \quad (2.21)$$

2.2.2 Three-Gage (0° 120° 240°) or (0° 60° 120°) Rosettes

These rosettes, called also delta rosettes, offer a very slight potential advantage over the 45° rectangular rosette. The only reason is due to the greatest possible angle from one another that in some cases could increment the accuracy. The relationships are similar to the other ones but more complex: the θ values are:

$$\begin{cases} \theta_a = 0^\circ \\ \theta_b = 120^\circ \\ \theta_c = 240^\circ \end{cases} \quad (2.22)$$

The previous relationships give:

$$\begin{cases} \varepsilon_a = \frac{\varepsilon_x + \varepsilon_y}{2} + \frac{\varepsilon_x - \varepsilon_y}{2} = \varepsilon_x \\ \varepsilon_b = \frac{\varepsilon_x + \varepsilon_y}{2} + \frac{\varepsilon_x - \varepsilon_y}{2}(-0,5) + \frac{\gamma_{xy}}{2} \cdot \left(-\frac{\sqrt{3}}{2}\right) \\ \varepsilon_c = \frac{\varepsilon_x + \varepsilon_y}{2} + \frac{\varepsilon_x - \varepsilon_y}{2}(-0,5) + \frac{\gamma_{xy}}{2} \cdot \frac{\sqrt{3}}{2} \end{cases}$$

$$\begin{cases} \varepsilon_x = \varepsilon_a \\ \frac{\gamma_{xy}}{2} = \frac{1}{\sqrt{3}}(\varepsilon_c - \varepsilon_b) \\ \varepsilon_y = \frac{1}{3} \cdot [2(\varepsilon_b + \varepsilon_c) - \varepsilon_a] \end{cases}$$

Following the same way as the 45° rosettes, the principal strains are:

$$\begin{cases} \varepsilon_1, \varepsilon_2 = \frac{\varepsilon_x + \varepsilon_y}{2} \pm \sqrt{\left(\frac{\varepsilon_x - \varepsilon_y}{2}\right)^2 + \left(\frac{\gamma_{xy}}{2}\right)^2} \\ \tan 2\phi = \frac{\gamma_{xy}}{\varepsilon_x - \varepsilon_y} \end{cases} \quad (2.23)$$

$$\begin{cases} \varepsilon_1, \varepsilon_2 = \frac{\varepsilon_a + \varepsilon_b + \varepsilon_c}{3} \pm \sqrt{\left(\frac{2\varepsilon_a - \varepsilon_b - \varepsilon_c}{3}\right)^2 + \frac{1}{3}(\varepsilon_c - \varepsilon_b)^2} \\ \tan 2\phi = \frac{\sqrt{3}(\varepsilon_c - \varepsilon_b)}{2\varepsilon_a - \varepsilon_b - \varepsilon_c} \end{cases} \quad (2.24)$$

ϕ is the angle between direction 1 and x axis. The principal stresses σ_1 and σ_2 are:

$$\begin{cases} \sigma_1, \sigma_2 = E \left[\frac{\varepsilon_a + \varepsilon_b + \varepsilon_c}{3(1-\nu)} \pm \frac{1}{1+\nu} \sqrt{\left(\frac{2\varepsilon_a - \varepsilon_b - \varepsilon_c}{3}\right)^2 + \frac{1}{3}(\varepsilon_c - \varepsilon_b)^2} \right] \\ \phi = \frac{1}{2} \arctan \frac{\sqrt{3}(\varepsilon_c - \varepsilon_b)}{2\varepsilon_a - \varepsilon_b - \varepsilon_c} \end{cases} \quad (2.25)$$

The first becomes:

$$= E \cdot \left\{ \frac{\varepsilon_a + \varepsilon_b + \varepsilon_c}{3(1-\nu)} \pm \frac{\sqrt{2}}{3(1+\nu)} \sqrt{(\varepsilon_a - \varepsilon_b)^2 + (\varepsilon_b - \varepsilon_c)^2 + (\varepsilon_c - \varepsilon_a)^2} \right\}$$

These formulas are valid for the sequence of angles:

$$\begin{cases} \theta_a = 0^\circ \\ \theta_b = 120^\circ \\ \theta_c = 240^\circ \end{cases} \quad (2.26)$$

Vice-versa, for the sequence:

$$\begin{cases} \theta_a = 0^\circ \\ \theta_b = 60^\circ \\ \theta_c = 120^\circ \end{cases} \quad (2.27)$$

a sign change is necessary in the second of Eqs. 2.25:

$$\phi = \frac{1}{2} \arctan \frac{\sqrt{3}(\varepsilon_b - \varepsilon_c)}{2\varepsilon_a - \varepsilon_b - \varepsilon_c} \quad (2.28)$$

2.3 Potentiometric Circuit

The strain gage is inserted in an electric circuit that transforms the measurement of the fractional electrical resistance of the grid into a measurement of fractional electrical voltage. The most simple circuit for measuring the fractional resistance of a gage is the potentiometric circuit, Fig. 2.8. The circuit is suitable only for measuring dynamic strains but it cannot be used for static or slowly variable deformations.

The output voltage at open terminals, when the strain gage is stretched, is:

$$e = \frac{R_1}{R_1 + R_2} \cdot e_0 \tag{2.29}$$

When the gage changes its resistance by the amount ΔR_1 , the output voltage is:

$$\frac{\Delta e}{e_0} = \frac{R_1 + \Delta R_1}{R_1 + R_2 + \Delta R_1} - \frac{R_1}{(R_1 + R_2)} = \frac{\Delta R_1 R_2}{(R_1 + R_2)^2 + \Delta R_1(R_1 + R_2)} \tag{2.30}$$

This is a highly non-linear relation with respect to ΔR_1 . The task of R_2 is to reduce the non-linearity and this happens if $(R_1 + R_2)$ is much greater than ΔR_1 . E.g. for $R_1 = R_2 = 120 \Omega$, the sum 240 is much greater than ΔR_1 variable in the range of 10^{-3} e 1Ω Only for very high deformation values must the non-linearity of the previous expression be taken into account.

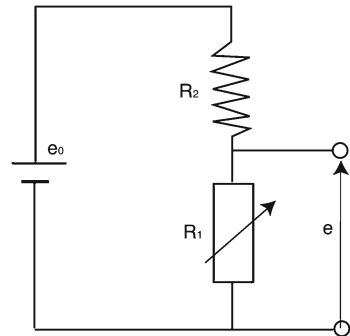
For the most common ΔR_1 values, $\Delta R_1(R_1 + R_2)$ is negligible with respect to the square of the sum and Eq. 2.30 gives with good approximation:

$$\frac{\Delta e}{e_0} = \frac{R_1 R_2}{(R_1 + R_2)^2} \cdot \frac{\Delta R_1}{R_1} = \frac{R_1 R_2}{(R_1 + R_2)^2} \cdot K \varepsilon \tag{2.31}$$

Since:

$$\lim_{R_2 \rightarrow 0} \frac{R_1 R_2}{(R_1 + R_2)^2} = \lim_{R_2 \rightarrow \infty} \frac{R_1 R_2}{(R_1 + R_2)^2} = 0 \tag{2.32}$$

Fig. 2.8 Scheme of the potentiometric circuit



high values of R_2 improve linearity but reduce the coefficient $\frac{R_1 R_2}{(R_1 + R_2)^2}$ that represents the circuit *sensitivity*; on the contrary, smaller R_2 values reduce both linearity and sensitivity.

For $R_2/R_1 = 1$ the circuit sensitivity reaches its maximum value equal to $1/4$.¹²

In this case:

$$\frac{\Delta e}{e_0} = \frac{K \varepsilon}{4} \quad (2.33)$$

The linearized approximation is obtained by the *first variation* of the functional e :

$$de = \frac{\partial e}{\partial R_1} dR_1 + \frac{\partial e}{\partial R_2} dR_2 + \frac{\partial e}{\partial e_0} de_0 \quad (2.34)$$

Second and third terms are zero if R_2 is a constant and if a stabilized excitation is used. Equation 2.34 becomes:

$$de = \frac{\partial e}{\partial R_1} dR_1 = \frac{R_2 e_0}{(R_1 + R_2)^2} \cdot dR_1$$

For finite differences, dividing and multiplying the second term by R_1 the fractional voltage change, proportional to the fractional resistance change is:

$$\frac{\Delta e}{e_0} = \frac{R_1 R_2}{(R_1 + R_2)^2} \cdot \frac{\Delta R_1}{R_1} = \frac{R_1 R_2}{(R_1 + R_2)^2} \cdot K \varepsilon \quad (2.35)$$

with the proportionality factor:

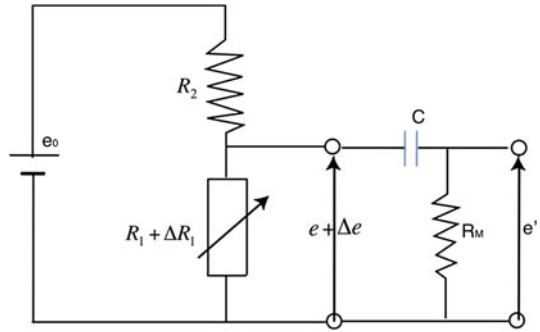
$$\frac{R_1 R_2}{(R_1 + R_2)^2} \cdot K$$

2.3.1 Measurement of Dynamic Strains

From the Eq. 2.31, applied to the case of classic strain values, the fractional voltage of the potentiometric circuit $\Delta e/e_0$ varies within the range $10^{-6} \div 10^{-2}$. With an excitation of a few volts, the minimum values of the voltage change Δe are expressed by the fifth or sixth decimal figure of the output voltage $e + \Delta e$. This problem can be overruled only for dynamic Δe measurement where it is possible to block the static component with a proper RC filter, Fig. 2.9.

¹² Numerator and denominator are derived by $(R_2)^2$: $\frac{R_1 R_2}{(R_1 + R_2)^2} = \frac{R_1/R_2}{(1 + R_1/R_2)^2}$; The maximum is obtained for the null value of the derivative with respect to the ratio $x = R_1/R_2$. The derivative $\frac{1-x}{(1+x)^3}$ is zero for $x = 1$, i.e. for $R_1 = R_2$.

Fig. 2.9 Potentiometer circuit with RC Filter



2.3.1.1 RC Circuit Response to an Impulsive Input

This problem is analyzed by J.F. Doyle, see [19] for a detailed presentation. Assuming a null internal resistance of the generator e , the output voltage is the drop $e' = R_M I$ across the resistor R_M (expression acceptable for open circuit or in the case of an output instrument with great resistance). The capacitor C stores energy such that the voltage drop is:

$$e_C = \frac{Q}{C}$$

from which:

$$\frac{de_C}{dt} = \frac{1}{C} \frac{dQ}{dt} = \frac{I}{C} \tag{2.36}$$

where: Q is the charge and I the current.

The input voltage e , i.e. the voltage drop on the series of capacitor and resistor R_M is:

$$e = e_C + e_R \quad \text{thus} \quad \frac{de}{dt} = \frac{de_C}{dt} + \frac{de_R}{dt} = \frac{I}{C} + \frac{de'}{dt}$$

where I is the current circulating in R_M that is equal to e'/R . Substituting Eq. 2.36, holds:

$$\frac{de}{dt} = \frac{e'}{RC} + \frac{de'}{dt} \tag{2.37}$$

or:

$$RC \frac{de'}{dt} + e' = RC \frac{de}{dt} \tag{2.38}$$

When RC is great, $e = e'$ and vice-versa, when RC tends to zero, then $e' = 0$. Thus, the potentiometric circuit can be used for recording dynamic phenomena, accepting inevitable attenuation and distortion effects, Fig. 2.10 (on the left side).

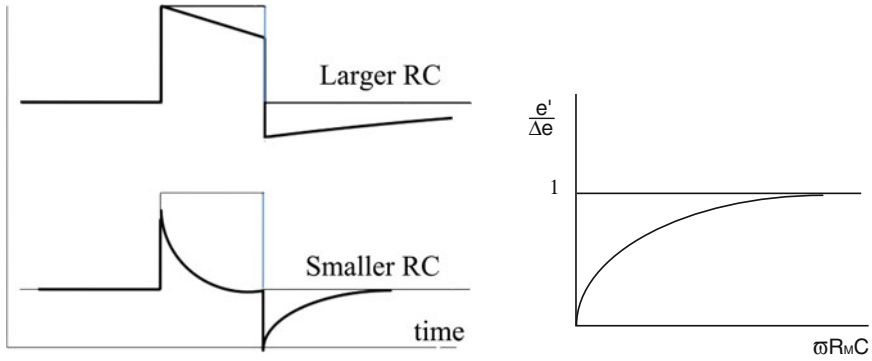


Fig. 2.10 Response of RC circuit to a rectangular [19] (on the left) and to a sinusoidal input (on the right)

2.3.1.2 RC Circuit Response to a Sinusoidal Input

Choosing a value of R_M much larger than the parallel of all the resistors before the filter, Fig. 2.9 (in such a way to neglect the voltage drops on the internal resistances of the supply voltage with respect to the filter impedance), the DC-current in R_M , due to a constant voltage e , is zero; the output voltage is due to the variable current only, generated by Δe , sinusoidal input.

In fact, taking the e expression, constituted by a constant term and a variable term: $e = e_0 + \Delta e \sin \omega t$, substituting this in the differential equation Eq. 2.38, gives:

$$RC \frac{de'}{dt} + e' = RC \cdot \Delta e \omega \cos \omega t \tag{2.39}$$

The term e disappears and the differential equation gives the following solution of the associated homogeneous Equation:

$$e_{omo} = K e^{-t/RC}$$

The diagram of the attenuation function, i.e. the filter characteristics is shown in Fig. 2.10 (on the right side)

The voltage drop e' across R_M terminals is then:

$$e' = R_M i_{var.} = \frac{R_M(\Delta e)}{|Z|}$$

Since the filter impedance is:

$$|Z| = \sqrt{R_M^2 + \frac{1}{\omega^2 C^2}} \tag{2.40}$$

finally holds:

$$\frac{e'}{\Delta e} = \frac{1}{\sqrt{1 + \frac{1}{(\omega R_M C)^2}}} \quad (2.41)$$

The constant component of the voltage is stopped while the variable component Δe is transmitted but attenuated as a function of R_M and C values.

Exercise 2.1 Calculate an RC filter for a dynamic measurement of a variable signal, proportional to a sinusoidal strain with an amplitude equal to $500 \mu\epsilon$ and a frequency of 30 Hz (corresponding to $\omega = 2\pi f = 190 \text{ rad/sec}$). The gage resistance R_1 is 120Ω , $K = 2$, and a grid area is $6 \cdot 2.8 \text{ mm}^2$. The resistor R_2 is equal to R_1 .

(a) Maximum Current in the strain gage

The limit value for P_D is 5 mW/mm^2 , (Thick steel). For the selected grid dimensions the limit power is $P_W = 84 \text{ mW}$. The corresponding maximum current is:

$$i = \sqrt{\frac{P_W}{R}} = 26.5 \text{ mA}$$

(b) Electromotive force (emf) and output voltage

If $R_1 = R_2$ we have:

$$i = \frac{e}{R_1} = \frac{e_0}{R_1 + R_2}$$

then:

$$e = 3.17 \text{ V} \quad e_0 = 6.35 \text{ V}$$

(c) Output voltage Δe

For $K = 2$ and $\epsilon = 1,000 \mu\epsilon$: $\Delta e = 1/4 K \epsilon \cdot e_0 = 2 \frac{10^{-3}}{4} \cdot 6.35 = 3.17 \times 10^{-3} \text{ V}$
The total output voltage is:

$$e + \Delta e = 3.17 + 3.17 \times 10^{-3} \text{ V}$$

(d) RC Filter

The resistance R_M must be as large as possible, to reduce output voltage drop:

$$R_M \gg \frac{R_1 R_2}{R_1 + R_2} = R_1/2 = 60 \Omega$$

If we select $R_M = 6,000 \Omega$, in order to transmit Δe with low attenuation, (e.g. $e' \geq 0.9 \cdot \Delta e$), it must be:

$$\frac{e'}{\Delta e} = \frac{1}{\sqrt{1 + \frac{1}{(\omega R_M C)^2}}} \geq 0.9$$

From this Equation the value of C is found:

$$C \geq 1.81 \times 10^{-6} \text{ Faraday} = 1.81 \mu F$$

2.4 Wheatstone Bridge

The constant output voltage component can be suppressed by two potentiometric circuits connected in opposition to each other as shown in Fig. 2.11. This circuit is called a Wheatstone Bridge. In this circuit the output voltage is only Δe , which can be measured on the principle of the null deflection method that is slow, but gives very accurate results. The other advantage of this circuit is its capability to perform logic operations such as signal products, sums and subtractions, that is a useful property for many applications. Because of its outstanding sensitivity, the Wheatstone bridge is the most frequently used circuit for strain measurements.

The output voltage of the circuit with four potentially variable resistances, can be written as the difference between two potentiometric circuits, Fig. 2.11 as:

$$V_C - V_A = R_3 i' - R_2 i = \left(\frac{R_3}{R_3 + R_4} - \frac{R_2}{R_1 + R_2} \right) \cdot e_0 \quad (2.42)$$

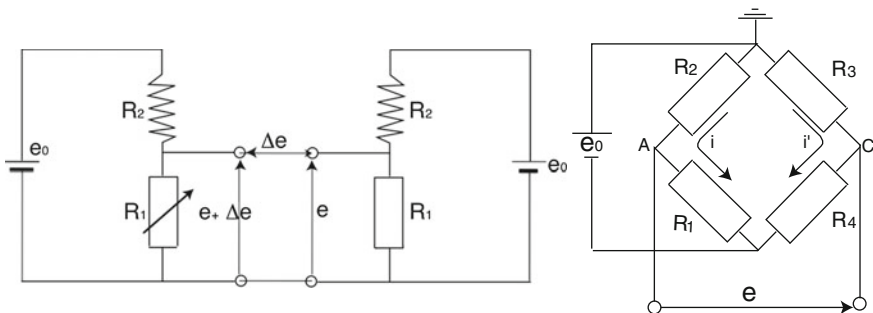


Fig. 2.11 Double potentiometric circuit in opposition and the equivalent Wheatstone bridge

An initial steady state voltage exists unless the numerator above is zero. Such a configuration with zero output voltage is termed a *Balanced Bridge* and is provided when:

$$\frac{R_3}{R_3 + R_4} = \frac{R_2}{R_1 + R_2}$$

i.e.:

$$R_1 R_3 = R_2 R_4 \quad (2.43)$$

If three of four resistances are known, the fourth (e.g. R_1) can be determined by the other values obtained at balance, without having any current on the output diagonal and without a stabilized electromotive force (emf) that does not play any role in the equilibrium equation.

If the bridge is used as a double potentiometer and the circuit gives an output produced by small changes in the resistance of the bridge arms, which can be considered infinitesimal changes, the circuit can be linearized and the following relationship holds:

$$de = \frac{\partial e}{\partial R_1} dR_1 + \frac{\partial e}{\partial R_2} dR_2 + \frac{\partial e}{\partial R_3} dR_3 + \frac{\partial e}{\partial R_4} dR_4 + \frac{\partial e}{\partial e_0} de_0 \quad (2.44)$$

Deriving first the infinitesimal changes, then substituting them with finite differences in each resistor:

$$\begin{aligned} \Delta e = & \frac{R_1 R_2 e_0}{(R_1 + R_2)^2} \cdot \left(\frac{\Delta R_1}{R_1} - \frac{\Delta R_2}{R_2} \right) + \frac{R_3 R_4 e_0}{(R_3 + R_4)^2} \cdot \left(\frac{\Delta R_3}{R_3} - \frac{\Delta R_4}{R_4} \right) \\ & + \left(\frac{R_3}{R_3 + R_4} - \frac{R_2}{R_1 + R_2} \right) \cdot \Delta e_0 \end{aligned} \quad (2.45)$$

The last term is zero if the bridge is balanced or for a stabilized supply voltage. The equation gives:

$$\frac{\Delta e}{e_0} = \frac{R_1 R_2}{(R_1 + R_2)^2} \cdot \left(\frac{\Delta R_1}{R_1} - \frac{\Delta R_2}{R_2} \right) + \frac{R_3 R_4}{(R_3 + R_4)^2} \cdot \left(\frac{\Delta R_3}{R_3} - \frac{\Delta R_4}{R_4} \right) \quad (2.46)$$

If the bridge is balanced before gage deformation, from Eq. 2.43 the following identities derive:

$$\frac{R_3}{R_3 + R_4} = \frac{R_2}{R_1 + R_2}$$

or:

$$\frac{R_4}{R_3 + R_4} = \frac{R_1}{R_1 + R_2}$$

multiplying the members of the first relation by the corresponding members of the second:

$$\frac{R_1 R_2}{(R_1 + R_2)^2} = \frac{R_3 R_4}{(R_3 + R_4)^2}$$

and Eq. 2.46 becomes:

$$\frac{\Delta e}{e_0} = \frac{R_1 R_2}{(R_1 + R_2)^2} \cdot \left(\frac{\Delta R_1}{R_1} - \frac{\Delta R_2}{R_2} + \frac{\Delta R_3}{R_3} - \frac{\Delta R_4}{R_4} \right) \quad (2.47)$$

Moreover, if $R_1 = R_2$:

$$\frac{\Delta e}{e_0} = \frac{1}{4} \cdot \left(\frac{\Delta R_1}{R_1} - \frac{\Delta R_2}{R_2} + \frac{\Delta R_3}{R_3} - \frac{\Delta R_4}{R_4} \right) \quad (2.48)$$

Remarks:

- Linearity simplification is acceptable only for small bridge unbalance voltage.
- Even if the bridge consists of four only nominally equal gages, Eq. 2.48 can be used considering four identical resistors (e.g. four strain gages of the same nominal values), because possible small variations on the real values of strain gage resistances do not affect the measurement accuracy. Even differences of 5 % due to the tolerance in the resistance of R_1 and R_2 produce errors less than 0.1 %.
- If the measuring points are limited number, it is right to balance the bridge before the test, to limit the initial unbalance currents.
- Vice-versa, in the case of Data Acquisition Systems based on a computer for the automatic scanning of a large number of strain gages, it is preferable to record and store the initial unbalanced of all the measuring points, subtracting automatically, at the end of the test, the initial values from the final readings.
- Commercial instrumentations offer many ways of balancing the bridge by series and parallel variable resistors, suitable also for balancing full bridges in which all the arms contain active strain gages, Fig. 2.12. In Fig. 2.13 the adjustable resistors are all in a second bridge in parallel with the main bridge, for a maximum balance operation flexibility.

2.4.1 Shunt Calibration

A substitution method is generally used when a direct method of calibration is not possible. In this method an auxiliary shunt is used that produces an effect on the measurement equipment comparable with a real predetermined mechanical strain. This is accomplished by shunting, or connecting, a large resistor of known value R_C across one arm R_1 of the bridge (in parallel with the nominal resistance), creating a known resistance variation equal to:

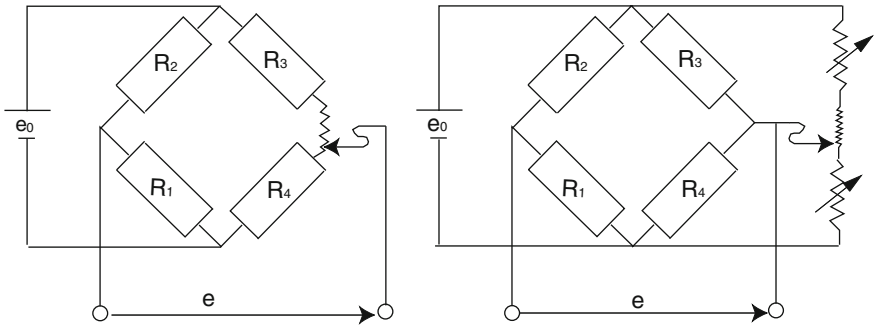


Fig. 2.12 Series and parallel circuits

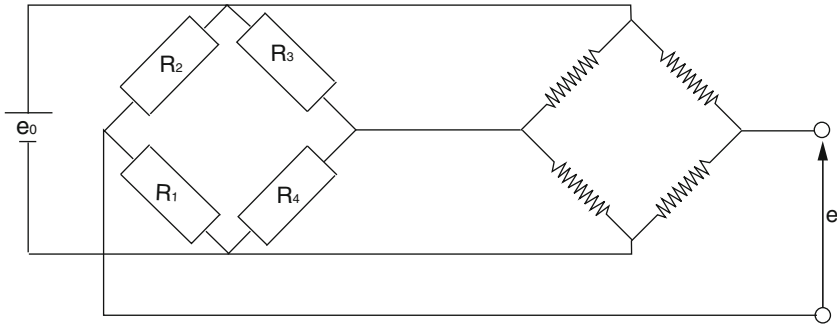


Fig. 2.13 A second bridge in parallel with the main one

$$\Delta R_1 = \frac{R_1 R_C}{R_1 + R_C} - R_1 < 0 \tag{2.49}$$

If the output of the bridge must correspond to a known deformation ε for strain gage with known gage Factor K , it is possible to write this condition:

$$\Delta R_1 = \frac{R_1 R_C}{R_1 + R_C} - R_1 = -K \cdot \varepsilon \cdot R_1 \tag{2.50}$$

that is negative since the parallel produces a reduction of the resistance of the arm 1. From this relationship the following shunt value derives:

$$\frac{R_C}{R_1} = \frac{1}{K \cdot \varepsilon} - 1 \tag{2.51}$$

Exercise 2.2 Calculate a Shunt for calibrating a circuit with: $K = 2$, $\varepsilon = 1,000 \mu\varepsilon$ and $R_1 = 120 \Omega$.

From Eq. 2.51

$$R_C = 59,880 \Omega$$

while, for $\varepsilon = 2,000 \mu\varepsilon$:

$$R_C = 29,880 \Omega$$

Since this choice is made with a conventional Gage Factor K equal $\hat{K}_{convent} = 2$, the actual deformation for the actual Gage factor K_{actual} , is:

$$\varepsilon_{actual} = \frac{2}{K_{actual}} \cdot \varepsilon$$

2.4.2 Bridge Excitation

The Wheatstone Bridge circuit can be excited by direct or alternating voltage [10]. The different choice is connected with the use of the amplifier that is necessary to raise the level of the signal from millivolts to volts.

- The direct voltage measuring amplifier provides a stabilized direct excitation of the bridge and amplifies static as well as dynamic signals also of high frequency. The disadvantage is the amplification, together with the main signal, of spurious signals due to electrical and magnetic field and thermal and galvanic voltages. The last can be detected by reversing the excitation.
- On the contrary, the alternating voltage measuring amplifier with the carrier frequency method amplifies static and dynamic signals with amplitude modulation of a high frequency carrier (225 Hz or 5 kHz), but does not accept frequencies above or below it. The spurious direct signals are eliminated but dynamic signals with a frequency greater than 9 Hz or 1 kHz respectively are cut. The alternating voltage also needs a double balancing device for resistance and capacitance balance, but is very stable when wires are properly protected against spurious capacitances.

The bridge equilibrium condition between the resistors and capacitors on the four arms (Fig. 2.14) is:

$$\mathbf{Z}_1 \mathbf{Z}_3 = \mathbf{Z}_2 \mathbf{Z}_4 \quad (2.52)$$

that originates two different relationships between moduli and phases:

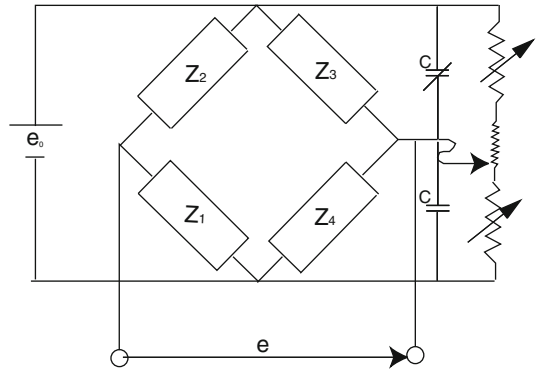
$$|Z_1| \cdot |Z_3| = |Z_2| \cdot |Z_4| \quad (2.53)$$

$$\phi_1 + \phi_3 = \phi_4 + \phi_2 \quad (2.54)$$

where:

$$|Z| = \sqrt{R^2 + \frac{1}{\omega^2 C^2}}$$

Fig. 2.14 Alternating voltage bridge



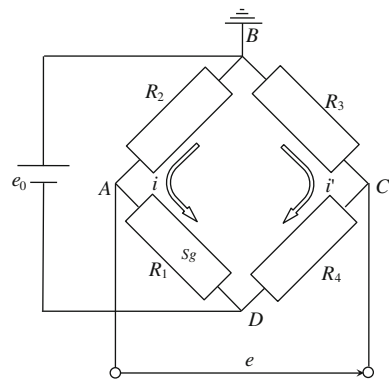
2.5 Bridge Configurations

The Wheatstone bridge offers the possibility to arrange a variety of circuits that, with computer aid, are able to perform accurate measurements at a huge number of points and to record and process data. Through the same kind of circuits, it is also possible to utilize strain gages devices for implementing control (see, e.g. in Chap. 8 the strain-controlled tests) and industrial surveillance systems.

2.5.1 Quarter Bridge

The most simple unit is the quarter bridge circuit for a single strain measurement. In this circuit three dummy resistances complete the bridge, as shown in Fig. 2.15. In this arrangement the output is temperature sensitive, unless the active gage (R_1) is self-compensated for that temperature range. The circuit is often used when accuracy

Fig. 2.15 Quarter bridge configuration



requirements are not severe ($\approx 10 \mu\epsilon$). The circuit is the same as the potentiometric one, and the output is given by Eq. 2.55:

$$\frac{\Delta e}{e_0} = \frac{R_1 R_2}{(R_1 + R_2)^2} \cdot \frac{\Delta R_1}{R_1} = \frac{R_1 R_2}{(R_1 + R_2)^2} \cdot K = \frac{K}{4} \epsilon_1 \quad (2.55)$$

(The last passage if $R_2/R_1 = 1$ and the circuit factor becomes 1/4).

2.5.2 Half Bridge

For this case two non-active resistances are provided in adjacent arms (R_2 and R_3), Fig. 2.16.

The bridge output is not sensitive to temperature so long as any temperature changes occur equally in the two arms R_1 and R_4 , provided both other elements are temperature insensitive. An output is produced only when unequal resistance changes are produced in gages R_1 and R_4 . The bridge output is:

$$\frac{\Delta e}{e_0} = \frac{R_1 R_2}{(R_1 + R_2)^2} \cdot K (\epsilon_1 - \epsilon_2) \quad (2.56)$$

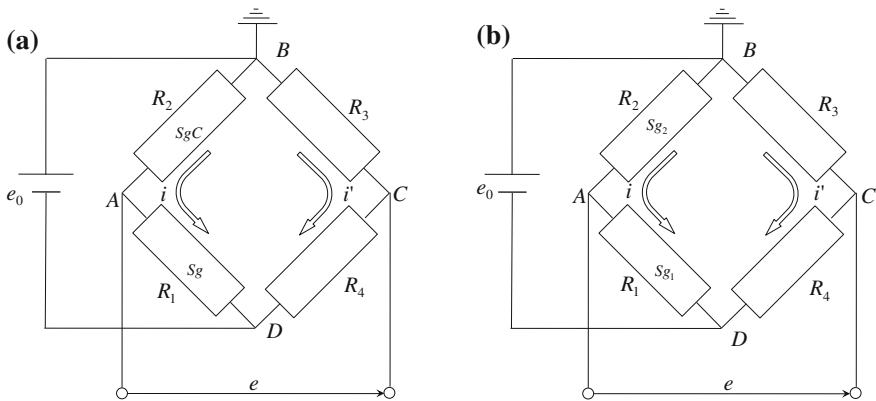


Fig. 2.16 Half bridge configuration with: **a** one active gage and a compensator gage, **b** two active gages

Two kinds of measurements are possible:

- If R_2 is a compensator gage, i.e. a non-active gage, with only an apparent deformation ($\varepsilon_2 = \varepsilon_{app}$), the output is:

$$\frac{\Delta e}{e_0} = \frac{R_1 R_2}{(R_1 + R_2)^2} \cdot K(\varepsilon_1 + \varepsilon_{app} - \varepsilon_{app}) = \frac{R_1 R_2}{(R_1 + R_2)^2} \cdot K \varepsilon_1 = \frac{\Delta e}{e_0} = \frac{K \varepsilon_1}{4} \quad (2.57)$$

If $R_2/R_1 = 1$, the last passage is valid.

- If the task of this circuit is the measurement of the bending strain in thin plates or beams, both the strain gages R_1 and R_2 are active and the output is twice that of a single active gage. If the two gages are close to each other, the possible temperature changes during the test will be nearly identical in them and no apparent strain will be observed. Moreover, if a component of uniform stretching (or compression) is present in addition to the bending, its effect would be canceled, since it would be the same for each gage.

The equation is the following, see Eq. 2.48 with $\varepsilon_1 = -\varepsilon_2$ and $\varepsilon_3 = \varepsilon_4 = 0$:

$$\frac{\Delta e}{e_0} = \frac{R_1 R_2}{(R_1 + R_2)^2} \cdot K(\varepsilon_1 - \varepsilon_2) = \frac{R_1 R_2}{(R_1 + R_2)^2} \cdot 2K \varepsilon_1 \quad (2.58)$$

This configuration has the following advantages in slender structure measurements:

- If the two gages are mounted opposite each other on opposite (top/bottom) surfaces so that a compressive strain ($-\varepsilon$) is measured by one and an equal tensile strain ($+\varepsilon$) by the other, Eq. 2.57 the output is twice the bending deformations (with an increment of the sensitivity).
- Apparent strains are eliminated.
- Possible contributions of strain due to normal load are also eliminated.

The same circuit but with the two active gages connected with opposite bridge arms cannot be considered a half bridge configuration since it does not eliminate apparent strains. Then, this circuit needs auto-compensated gages. Nevertheless, this circuit has the following advantage: it eliminates the bending deformations and gives double the normal strain value, improving the sensitivity of the measuring circuit.

2.5.3 Full Bridge

In a full bridge measuring circuit all the gages on the four arms are active. This circuit has perfect thermal compensation and maximum sensitivity.

The circuit equations are given by Eqs. 2.46–2.48 depending on the values of the four resistances in the bridge arms. The last one becomes:

$$\frac{\Delta e}{e_0} = \frac{K}{4} \cdot (\varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4) \quad (2.59)$$

Some applications are described below.

2.5.3.1 Axial Strain in Slender Beams

Two strain gages are mounted on two opposite sides of the bar, along the longitudinal direction and connected on opposite branches of the bridge in order to double the strain of the bar due to a normal load. The two additional gages are mounted on the same opposite sides of the bar but on transversal directions, Fig. 2.17 and connected on the remaining branches of the bridge. Through this scheme the following results are obtained and the Equation is the following, Eq. 2.60:

- The strains due to a bending moment M_y are not recorded.
- All the apparent strains are eliminated.
- The strains due to a bending moment M_x are also not recorded.

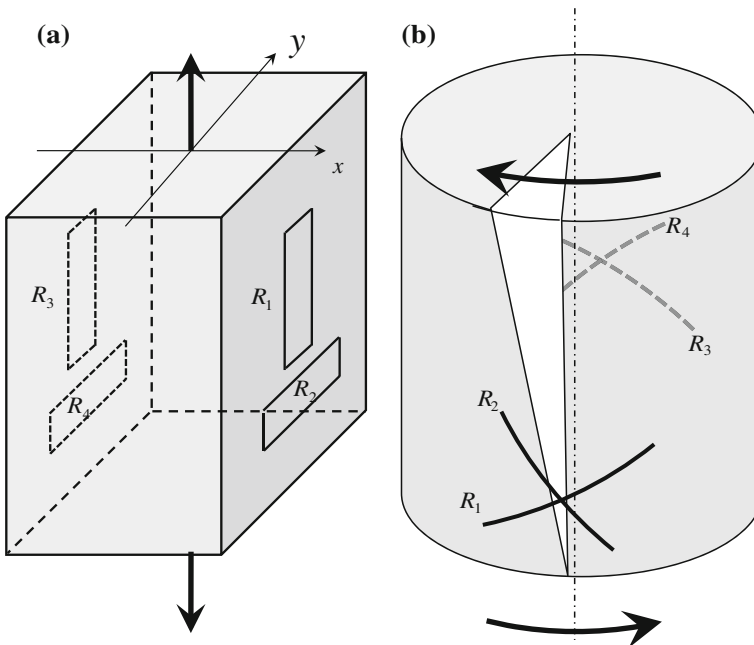


Fig. 2.17 Full bridge applications. **a** Measurement of axial stress in a bar; **b** measurement of torsion in a shaft

- The circuit sensitivity is improved due to the concurrent contributions of the four gages.
- The output signal is linearly related the axial load only.

$$\frac{\Delta e}{e_0} = \frac{K}{4} \cdot (\varepsilon_1 + \nu\varepsilon_1 + \varepsilon_1 + \nu\varepsilon_1) = \frac{2(1 + \nu)K}{4} \cdot \varepsilon_1 \quad (2.60)$$

2.5.3.2 Deformation of a Circular Shaft Subjected to Pure Torsion

The same electrical connection is utilized for measuring deformation of a circular shaft subjected to torsion and other spurious loadings such as bending moments.

To isolate the pure torsion, the gages on opposite sides of the bridge must be bonded on opposite parts of the shaft, but in the same helical direction. The general equation is:

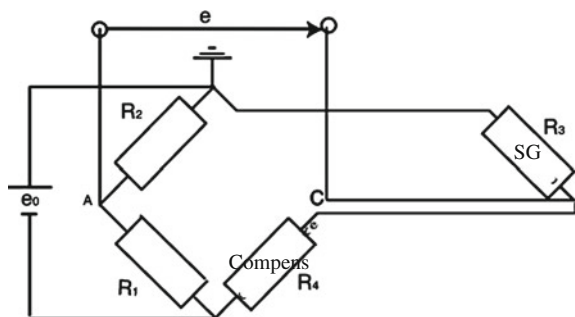
$$\frac{\Delta e}{e_0} = \frac{K}{4} \cdot (\varepsilon_1 + \varepsilon_1 + \varepsilon_1 + \varepsilon_1) = \frac{4K}{4} \cdot \varepsilon_1 = K\varepsilon_1 \quad (2.61)$$

Since all the strain gages measure the same value of strain, sensitivity of the full bridge is four times the sensitivity of one strain gage.

2.5.4 Eliminating Cable Effects with Three-Wire Circuit

For compensating thermal resistance variation of the wires when the strain gage is mounted far from the recording unit, the three-wire circuit is suggested [20]. Following the circuit of Fig. 2.18 it is possible to put in series a wire of the same length, at the active strain gage and at the compensator gage as well, canceling the thermal effect of the wire from the bridge output. In the Bulletin [20] the following advantages are summarized: *In summary, benefits of the three-wire circuit include intrinsic bridge balance, automatic compensation for the effects of lead-wire temperature changes*

Fig. 2.18 Three-wire circuit for eliminating thermal effect of wire resistance contribution



on bridge balance, and increased measurement sensitivity compared to the two-wire configuration. The three-wire hookup is the recommended configuration for quarter-bridge strain gage circuits for static strain measurement.

2.6 Semiconductor Strain Gage

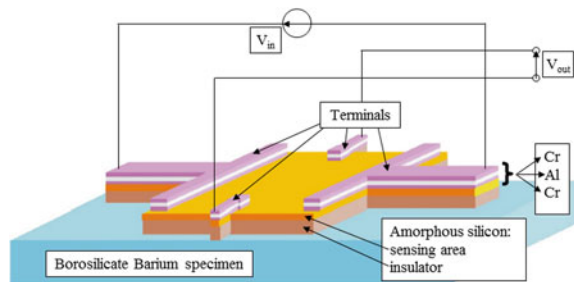
Some semiconductor materials, such as germanium and silicon, offer a strong piezo-resistive effect and special semiconductor gages can be made with a high nominal resistance and with small dimensions, with a range of application from 1,000 to 10,000 $\mu\epsilon$, and with either positive or negative resistance changes. The high output of the semiconductor gage can be efficiently used for measuring dynamic strains, where temperature effects are generally less important than for static strain.

The special range of application consists in the use in feedback loops of the force control of finger sensors of robots, where manipulated objects could be damaged due to too high applied force or could be dropped because of poor clamping. Possible applications include medical devices for non-invasive pressure measurements, machine tools, such as *CNC-lathes*, for continuous monitoring of the tool trajectory avoiding the undesired torsion moment at the cutting edge, the construction of miniaturized cells, etc. The advantage of a high sensitivity to strain is penalized by a non-linearity of change in resistance with strain.

2.7 Tests on Piezo-Resistive Hydrogenated Amorphous Silicon Strain Gages

A sensor that might overcome some of the limitations and drawbacks of traditional electrical strain gages, together with the need for an adhesive [21], can be developed on the basis of the piezo-resistive effect. The sensor active area, Fig. 2.19, consists of a thin n-doped hydrogenated amorphous silicon layer (a-Si:H).

Fig. 2.19 Sensor structure



Output voltage, representing the sensor response, greatly depends on the resistivity variation, whereas there is no dependence on sensor geometric strain or distortion. Some previous experiments [22] were able to show that the output signal, acquired at a perpendicular direction with respect to current density vector J , depends on the relative orientation between the bias current direction and the stress/strain principal coordinate system, according to the relationship:

$$V_{\perp} = \rho \pi_s \sigma_p |J| \quad (2.62)$$

where ρ refers to the resistivity of the amorphous silicon/silicide layers, π_s to the piezo coefficient and σ_p to principal stress. The dependency is due to π_s : it assumes zero value when J is parallel or perpendicular to the principal directions and maximum value, as J is $\pm 45^\circ$ inclined.

From an electrical point of view, the sensor can be compared to a full Wheatstone bridge with two input and two output terminals. While in a full bridge, gage resistances are located on each side arm, in this sensor the resistance is distributed over the whole sensing area and the resistances between the contacts are due to surface conductivity of the amorphous film. The sensor can be powered in current or voltage, e.g. a supply voltage of 2.5 V is applied.

2.7.1 Gage Preparation and Testing

The amorphous silicon material is deposited on a glass specimen by *PECVD* (plasma-enhanced chemical vapor deposition at high vacuum pressure) in a stainless steel chamber, Fig. 2.20. Tetraethoxysilane gas was used to transfer silicon before its deposition. Deposition data are the following:

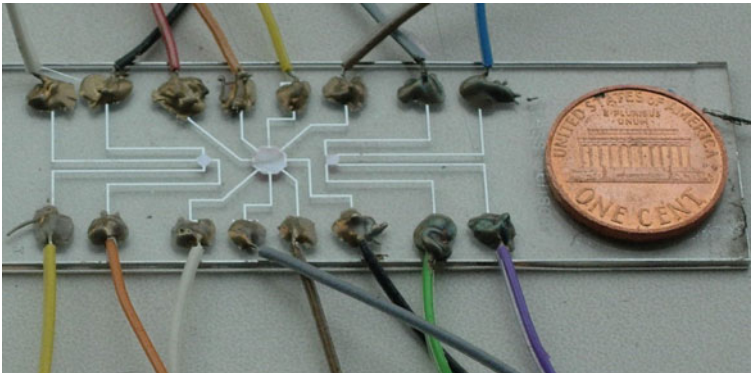


Fig. 2.20 Sensors deposited on a glass specimen: dimensions compared to one dollar cent coin

- Gas flow of pure silane 40 sccm.
- Silane diluted phosphine 10 sccm.
- Pressure 40 Pa.
- Power density 40 mW/cm².
- Deposition temperature 150 °C.
- Thickness 500 nm.

Three sensors with different shapes and dimensions, applied to a glass specimen, are considered. One of them has an octagonal sensitive area measuring 4.83 mm², with a side dimension of 1 mm and a thickness of 10 μm. Figure 2.21 shows all the configurations tested for each sensor.

1. The octagonal sensor can be supplied along a 45° direction 5–6 and the output voltage is read at the perpendicular direction 15–8 (configuration *A*). Otherwise, it is possible to invert the input voltage with the output one (configuration *B*). Other options: supplying and reading the output along vertical and horizontal directions, (configuration *C* and *D*).
2. The second sensor has a square shape with reduced dimensions. The sensitive area is 0.09 mm² (side 0.3 mm long), but input and output direction can only be 45° inclined, (configurations *E* and *F*).
3. A third sensor with a rhomboid shape has similar dimensional characteristics to the square one, but voltage input and output reading can be performed only along horizontal and vertical directions (configurations *G* and *H*).

The resistance related to the distributed conductivity of the sensing area is in the order of 7 kΩ, a value being ten times higher than that of conventional strain gages. Power consumption is therefore much lower, around 1 mW.

The response of the piezo-resistive sensor was compared to that of two conventional strain gages in the longitudinal and transverse directions under pure bending load, by the device in Fig. 2.22.

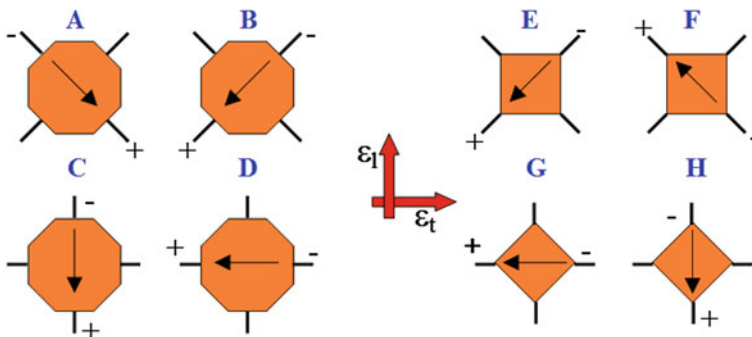


Fig. 2.21 Sensor configurations, where ε_l and ε_t are the longitudinal and transverse strains

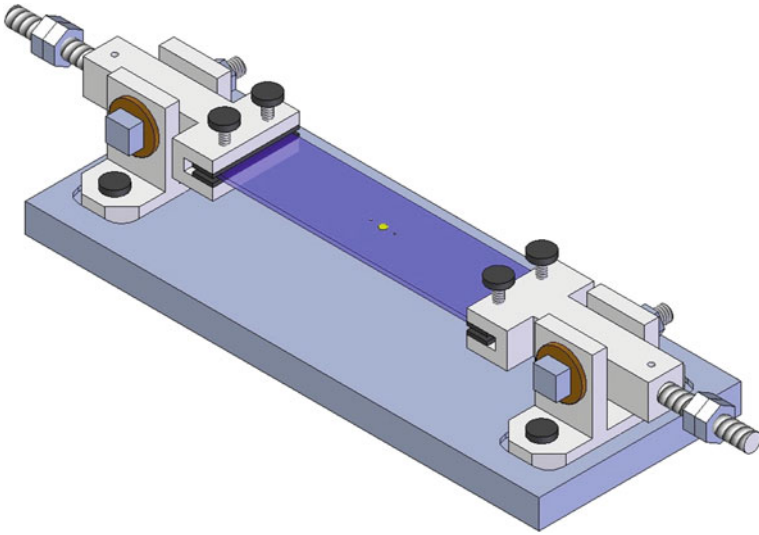


Fig. 2.22 Testing device for pure bending moment application

2.7.2 Calibration and Test Results

The results are shown in the following figures and compared with strains measured by standard gages. All the configurations described in the previous paragraph were investigated. Only some results are presented here (for a complete description see [21]) and plotted in Fig. 2.23 for configuration *A* and in Fig. 2.24 for configuration *E*. The experimental response shows that output voltage is strictly related to the minimum angle between the supply direction and one of the principal directions. In the cases of configurations *A* and *B*, Figs. 2.23 and 2.24, the voltage supply

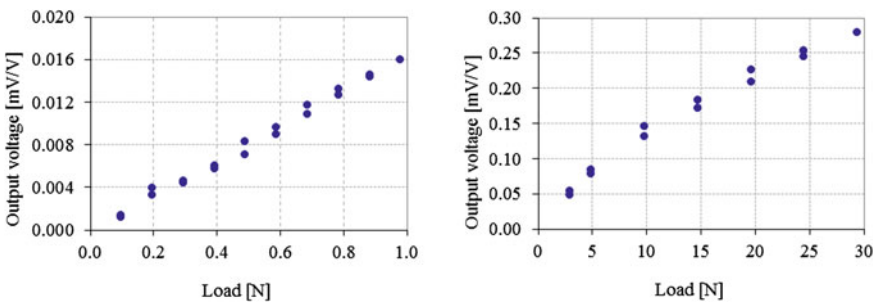


Fig. 2.23 Output voltage of octagonal shape sensor with configuration *A*: low (on the *left*) and high force ranges (on the *right*)

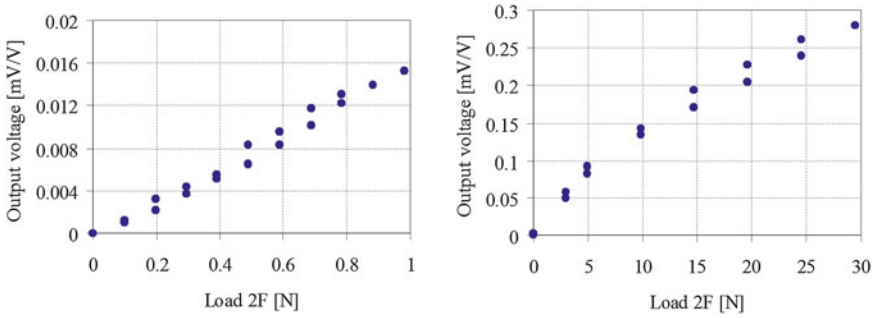


Fig. 2.24 Output voltage of square shape sensor with configuration *E*: low (on the *left*) and high force ranges (on the *right*)

direction is 45° inclined with respect to both principal directions. The same occurs for configurations *E* and *F*.

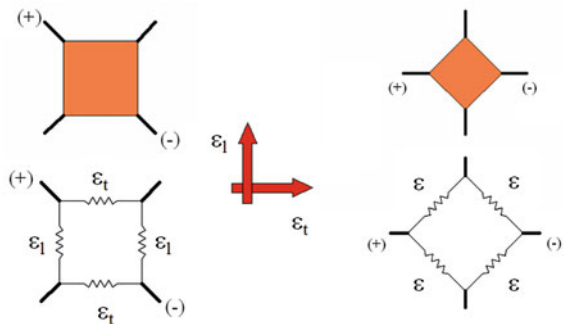
In order to evaluate sensor linearity and sensitivity, a comparison between different configurations of piezo-resistive sensors and Wheatstone bridges can be done. A full bridge can be associated with every configuration with the condition that supply and output directions are the same. Under this hypothesis the four arms are ideally located along the edges of the piezo-resistive sensor shape.

In the case of the square shaped sensor (configurations *A* and *B*, Fig. 2.21 with 45° inclined supply direction), the equivalent bridge is indicated in Fig. 2.25 (on the left), where ϵ_l and ϵ_t respectively are the longitudinal and the transverse strain of the calibration bar.

In the case of the rhomboid configuration (*G* and *H*) the equivalent full bridge is shown on the right. All the equivalent strain gages on the bridge arms withstand the same strain (ϵ) at 45° inclined directions with respect to the sample axis.

If the strain gages are ideally located on the edges, two of them would be parallel to the specimen axis and the other two would have transverse orientation. By using basic rules of the Theory of Elasticity, the bridge output voltage equation can be computed as follows, Eq. 2.63.

Fig. 2.25 Interpreting performance of square and rhomboid sensors as equivalent Wheatstone bridges



$$\frac{\Delta e}{e_0} = \frac{1}{4} K \cdot 2(1 + \nu) \varepsilon_l \tag{2.63}$$

where ε_l is the strain along the longitudinal axis of the calibration specimen. The square sensor output voltage can be related to the term $2(1 + \nu) \varepsilon_l$, estimated by using ε_l , and ν measured by resistance strain gages. In this way a *Gain Coefficient*, equivalent to the resistance *Gage Factor*, is given in Eq. 2.64.

$$K = \frac{\Delta e}{e_0} \frac{2}{(1 + \nu) \varepsilon_1} \tag{2.64}$$

The relationships and regression straight lines between output voltage and strains are shown in Fig. 2.26, with reference to configuration A. In the case of the rhomboid sensor (configurations G and H), the strain gages, ideally oriented along 45° inclined directions, withstand the same strain. Therefore, the output voltage is zero.

The Gain factor K from Eq. 2.64 is proportional to the slope of the linear relationship between output voltage and bridge reading. The interpolation suggests that K is not influenced by the load intensity and is equal to one for all configurations.

2.7.3 Temperature Response

The described piezo-resistive sensors were also tested to determine their response under temperature variation in different load conditions. The mean values of voltage offsets per unit variation of temperature stand in the range from 0.013 mV/°C (configuration F) to 0.24 mV/°C (configuration H), with a mean value of 0.12 mV/°C. The described results can be compared to other outcomes in the MEMS literature. A similar experimental campaign was performed by Pramanik et al. [23], considering a piezo-resistive porous silicon pressure sensor. The retrieved mean value of voltage offset per unit variation of temperature is about 0.15 mV/°C, which appears to

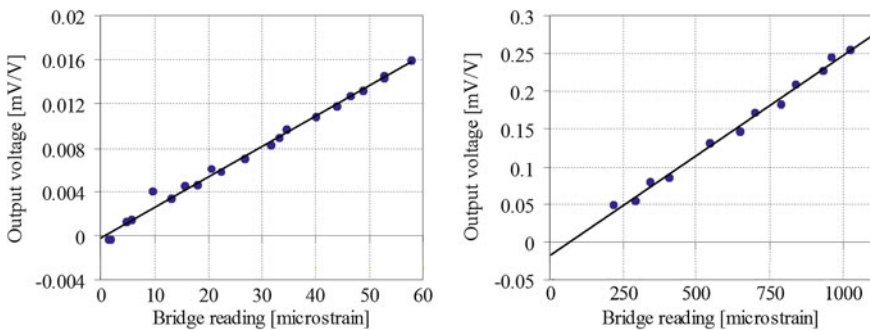


Fig. 2.26 Configuration A sensor interpreted as a Wheatstone bridge

be comparable to the previously mentioned value ($0.12 \text{ mV}/^\circ\text{C}$) for the described piezo-resistive sensors. For full details see [24].

2.8 Load Cells with Strain Gages Sensors

The following sections deal first with some theoretical issues on one and multiple-degree of freedom load cells. Afterwards, the manufacturing trends, the calibration procedure and the results are presented with reference to different degree of freedom load cells.

Load Cells are the essential components of every modern force measuring system in universal machines as well as in special laboratory equipment. The adoption of load cells is often the only option to reliably estimate actual loads in operation conditions by *in-field* or lab tests by specifically developed devices. Many examples of design and applications of custom cells are reported in [21, 25–40]. A further interesting trend stands in the development of identification techniques, for the direct determination of mechanical properties. An accurate measurement of strain is the first step for the application of these techniques [41–45].

Load cells consist of an elastic element which must be deformed under the action of the force linearly to the applied load. They are made of material that must have linear elastic behavior, no hysteresis and protection from the environment, such as hardened, stainless steels, heat treated aluminum alloys and copper-beryllium alloys [46].

The elastic element is instrumented with strain gages with resistance generally of 350Ω and epoxy adhesive, arranged in a full bridge, with the addition of calibration and compensation resistors. The main features of a load cell can be summarized in the following.

- It must be designed for suitable measurement of a minimum load with a required accuracy: in particular, the output voltage must be measurable even at the lowest load corresponding to cell resolution.
- At the same time, the cell strength must be warranted when the maximum load is applied, considering a safety coefficient of at least two.
- A cell must be design to fulfill the requirements concerning the size of the instrumented part and the overall dimensions.
- The output voltage and consequently the yield measurement should be independent of unexpected effects, such as a not precise force application point, force inclination generating bending, etc.

2.8.1 Bending Load Cells

Like elastic elements, Bending Load Cells have a cantilever beam in bending, in which the signal due to deformation is amplified by the lever effect and by the mounting of two strain gages on opposite sides of the bar (see Sect. 2.5.2). These cells are suitable for low load measurement, well compensated for temperature variation. The solution of Fig. 2.27b has a tapered shape of the instrumented section to offer constant bending strain in that part, eliminating the gradient of deformation in the grids of the strain gages (a beam with constant section, Fig. 2.27a has a triangular profile of bending strain). The reduction of the beam mass increases the natural frequencies. Two drawbacks are due to deflection that causes a beam length, (thus, a bending moment) reduction and a normal stress component, both origins of a non-linearity.

For these reasons, loading cells are developed according to the static scheme of Fig. 2.28 and designed as in Fig. 2.29. The strain gages are mounted at full-bridge by placing those with strains of the same sign on the opposite sides of the bridge. The tensile strain component in every beam, present in the four strain gages, is automatically compensated for the imposed deformation due to rigid constraints.

2.8.2 Ring Load Cells

A ring was the first cell used for force measurements, Fig. 2.30; it was suitable as a mechanical load cell if equipped with an internal instrument which measured

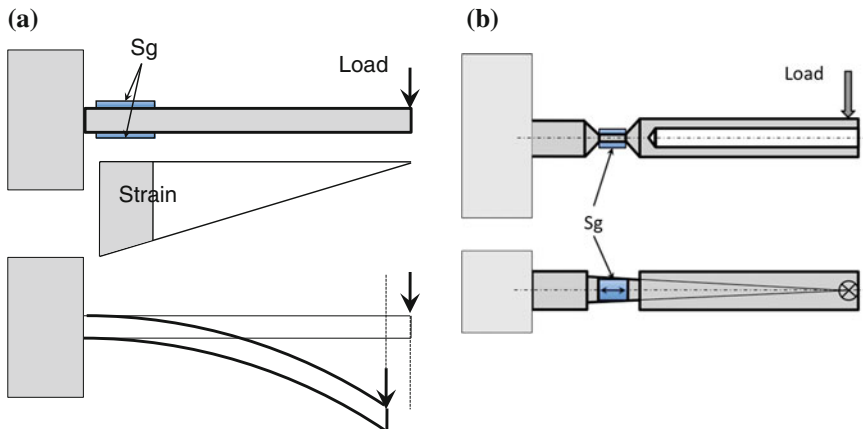


Fig. 2.27 The simplest example of the bending cell. **a** A cantilever beam with a load at one end, instrumented with two opposing strain gages and **b** an improved development of the basic idea by reducing the thickness and tapering the measuring area in order to make the strains in the gage grid uniform

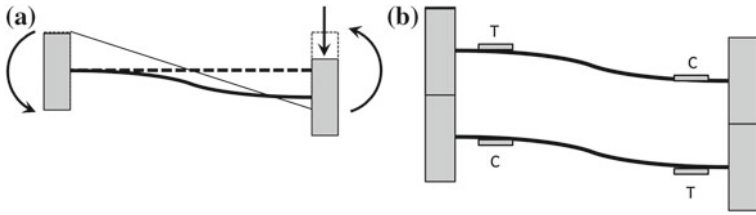


Fig. 2.28 Static schemes of load cells with single beam: **a** doubly fixed, with loads and moments applied to the opposite ends and **b** with double beam, instrumented with four strain gages on the opposite sides

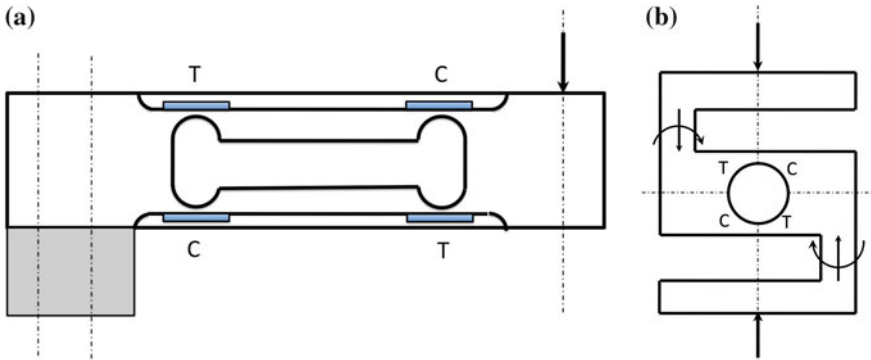


Fig. 2.29 Conceptual schemes of load cells with double bending beams **(a)**, (with a risk of variations on the load direction) and with double bending beams degenerated into a central hole of a thick block **(b)**, with a load that keeps its alignment more easily

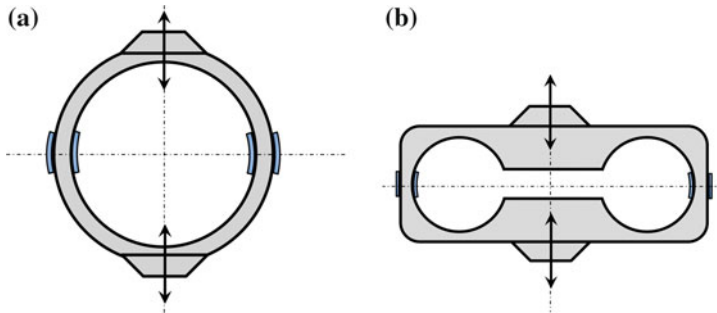


Fig. 2.30 Schemes of ring cells with circular **a** and oval **b** shape

the diameter deformation when the load was applied. It is still marketed for its simplicity, accuracy and precision, instrumented by four strain gages to form a full-bridge. Fitted with four electrical strain gages, it is an excellent load cell that can be optimized to reduce the sensitivity to slightly misaligned loads and to offer a wide range of applications and load capacities. Four strain gages are mounted on

the diametral plane perpendicular to the loads, inside and outside the ring. The traction components, which are identical in the four strain gages, are compensated and removed from the measurements, while the bridge detects four times the single strain of bending of one gage. The oval solution (*b*) has a low dimension in one direction [8].

2.8.3 Shear Load Cells

An important feature of load cells, especially for fatigue tests, is the stiffness, that must be as high as possible in order to avoid unwanted deformations and variations of the vibration modes of the specimen.

The dynamic part of the testing machines based on this principle [47], mainly consists of a series of masses, plus the specimen mass, and all elasticities of the specimen and of the other parts within the dynamic load flow. These parts form the oscillating system which is controlled and excited in its resonant frequency.

A low stiffness could reduce the fatigue life of the cell itself and vary the test duration. Besides, with millions or tens of millions of cycles, the drawback could be very serious. Material and structure-testings needs high stiffness cells but sensitive to small deformations.

A type that meets this criterion works on the principle of measuring the shear strains, Figs. 2.31 and 2.32. The local deformation of a small area, with proper dimensions of its thickness, can be sufficiently high to allow measurement with good sensitivity, while the global deformation (i.e. the displacement of the points of the load application) remains low. Unlike the bending cells in which the flexural deformations depend on the geometry variation during the arm deformation, thus by the displacement of the points of load application, the shear stresses are directly

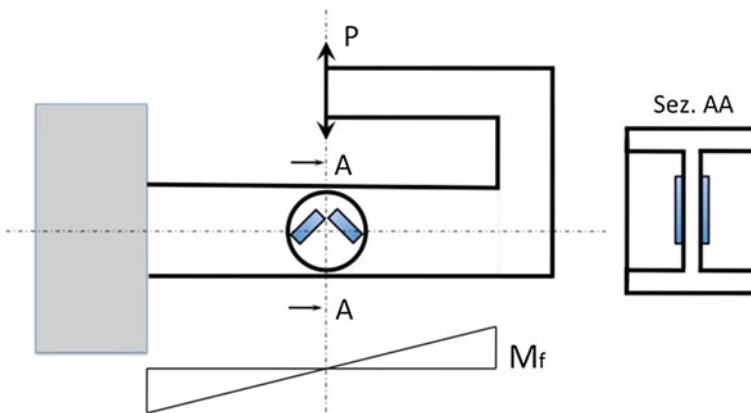
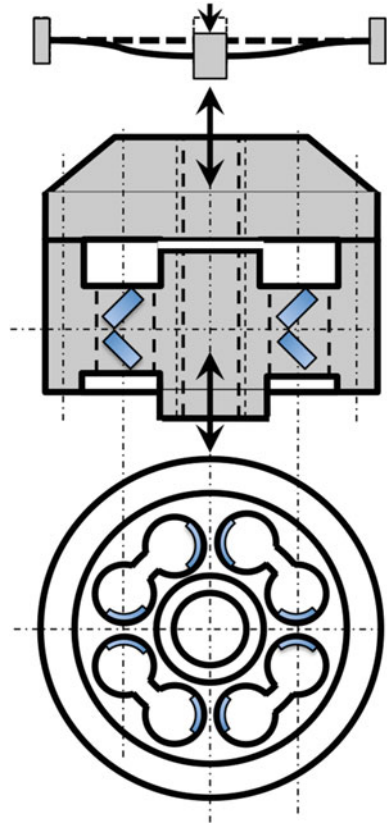


Fig. 2.31 Scheme of an elementary shear cell

Fig. 2.32 Sketch of a cell with high stiffness: the massive part guarantees that the deformed circular plate has a perfect constraint at the external circumference



related only to the load (regardless of point the of application) and to the body thickness. Varying the thickness will directly vary the shear stresses. Limited spaces can be machined in a body, large enough to accommodate the strain gages, but with a thin wall thickness, to produce high shear deformations.¹³ In this way, cells of high sensitivity, stiffness and load capacity can be built.

A final comment. Shear stresses are always accompanied by bending stresses ($T = dM/dx$) which, in this case, represent a disturbance for the measurement. It is therefore advisable to study a cells shape that allows shear stress measurement in areas with minimum values of the bending moment. E.g., in the scheme of Fig. 2.31 the measurement of shear deformation is confined to a point, with virtually zero bending moment.

With the same concept, the load cell of Fig. 2.32, suitable for fatigue testing machines, was developed. This cell is not very sensitive to misalignment of the load, thanks to a wide possibility of circuit compensation due to links with two bridges. The cell is developed for a load capacity from 2.5 to 5 kN and beyond; it consists of a

¹³ With an order of magnitude of $1,000 \div 1,700 \mu\epsilon$, with output bridge voltages of $2 \div 3 \text{ mV/V}$.

circular body with eight holes, with gages bonded at the points of thin walls to detect high shear strains. This circular element is connected through a series of screws to a massive structure that has the function of stiffening the instrumented area, limiting the bending deformation and enhancing the shear deformation.

2.9 Load Cell Calibration

Theoretical formulations could be used in the definition of the relationship between the applied load and the resulting strains.

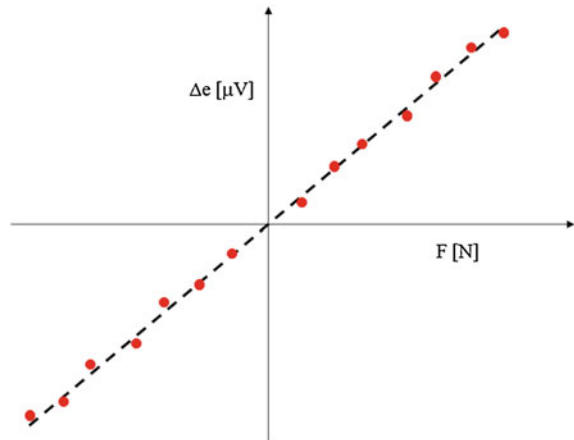
However, systematic errors could enter, due e.g. to positioning errors (on location and on direction) of strain gages, to the average effect of the measured strain on the grid length, and to the grid cross sensitivity, etc. For all these reasons it is always necessary to perform an experimental calibration.

A quite rigid calibration protocol must be followed. The step by step application of increasing and decreasing loads is recommended. In no-loading condition the zero output is checked and, if necessary, adjusted. Once a step has been established, the applied force (or moment) is incremented and the resulting output voltage is retrieved and recorded. This procedure is repeated until the force (or moment) value at full scale is reached. It is sometimes advisable to overcome this value by about 10 or 15 %. This procedure is usually followed to make sure of a complete adhesion between the applied strain gage and the instrumented surface. Acting in this way improves the accuracy and the repeatability of measurements.

Afterwards, load is step by step decremented to zero, with data recording at each step. When the cell is designed to measure both positive and negative loads, the direction of the force (or moment) is inverted and its value is again incremented up to full scale or to a 10 ÷ 15 % more. The resulting output voltage is measured at each loading step. Afterwards, load entity is decreased to zero, with data recording at each step. The entire procedure must be repeated at least three times. Finally, a diagram relating the voltage output to the applied load is determined.

The experimental data are usually linearly interpolated to obtain the empiric relationship between the force, indicated as F , and the output voltage Δe , Fig. 2.33. Once the calibration formula is known, the cell can be used to retrieve the unknown force, based on the measured output voltage.

Fig. 2.33 Typical linear response of a calibrated load cell: relationship between the output voltage and the applied load



The functioning logic of a loading cell can therefore be summarized in the following points.

- As load is applied, strains are generated on the instrumented elastic element of the cell.
- Consequently, strain gages experience variations of their nominal electrical resistances.
- These variations are detected by measuring the output voltage of a Wheatstone bridge.
- From the knowledge of the force—output voltage ($F - \Delta e$) analytical relationship, the entity of the applied load can be determined.

2.10 Load Cells with More Degrees of Freedom

When there is interest in the simultaneous measurement of more than one loading component, cells with more degrees of freedom can be used. They offer the advantage of simultaneous detection of several or all loading components: this is necessary for fatigue tests in case of multiple loads acting simultaneously. The strain gages are connected to form as many bridges as the forces or moments to be measured. In other words, as many channels as unknown terms are necessary.

A multiple-degrees load cell should be decoupled, i.e. each channel must be highly sensitive to just one force (or moment) component and theoretically insensitive to all the other ones.

The load cells are essentially of two types, see Chap. 1:

- Isostatic cells, in which the six components of force and moment are measured by the same number of independent units and the inverse problem is statically determined.
- Statically indeterminate cells, in which the number of measurements is greater than the number of unknown variables and the inverse problem is overdetermined.

The first are simpler and do not suffer from ill-conditioning of the measurements, while the latter are more complex, but can offer advantages related to the mechanical safety of the cell and to the construction compactness.

The most general case is that of a *six-degree of freedom* load cell: this cell type is able to measure three force and three moment components.

Let $[F_x, F_y, F_z]$ be the vector of the force components in xyz reference systems and let $[M_x, M_y, M_z]$ be the vector of the moment components in the same directions. These two vectors can be merged in a six-sized vector containing all the unknowns \mathbf{f}_i , with $i = 1, 2, \dots, 6$.

Measurements are fulfilled by processing the output voltages Δe_i of the six channels, collected in the vector Δe_i , with $i = 1, 2, \dots, 6$. The relationship between these vectors can be expressed by a 6×6 matrix, as in Eq. 2.65.

$$\Delta \mathbf{e} = \mathbf{B} \cdot \mathbf{f} \quad (2.65)$$

$$\begin{pmatrix} \Delta e_1 \\ \Delta e_2 \\ \Delta e_3 \\ \Delta e_4 \\ \Delta e_5 \\ \Delta e_6 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} & b_{16} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} & b_{26} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} & b_{36} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} & b_{46} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} & b_{56} \\ b_{61} & b_{62} & b_{63} & b_{64} & b_{65} & b_{66} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \end{pmatrix}$$

The term \mathbf{B} is usually regarded as the *Compliance Matrix* and has as many rows as the number of channels and as many columns as the number of degrees of freedom. For a six-d.o.f. load cell, it is typically a 6×6 matrix.

If the transducer is *decoupled*, the terms on the principal diagonal are theoretically zeroes and the matrix becomes diagonal. In actual conditions, it implies the terms on the diagonal being much greater than the other ones. The calibration consists in the determination of the terms b_{ij} of the matrix \mathbf{B} .

The generic term b_{ij} stands for the output voltage of the i -th channel (Δe_i) when the j -th component (F_j) is unitarian and the all the other ones are zeroes. The consequence is that the calibration procedure must be conducted with a separate application of all the load components. The protocol to be followed is similar to that described for the one-degree of freedom load cells, with step by step increasing and decreasing and output recording: at each step all the six outputs must be saved.

Considering a load cell with a good decoupling, the data after the application of the third load component are qualitatively shown in Fig. 2.34.

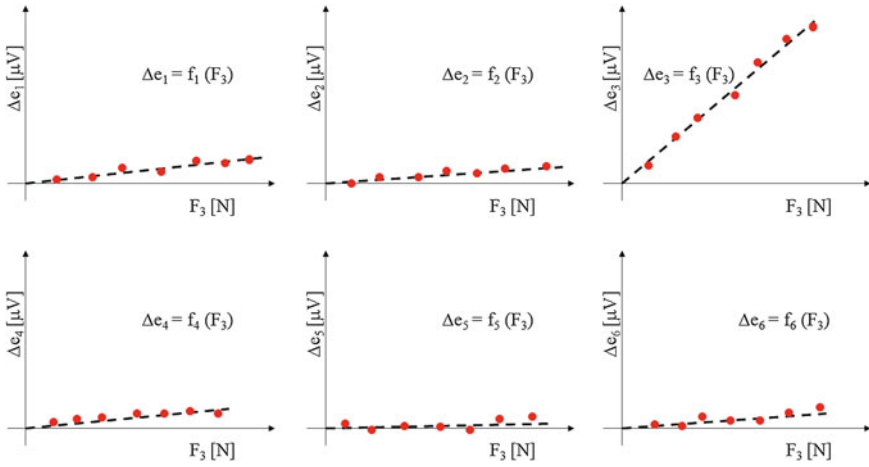


Fig. 2.34 Calibration procedure of a six-degree of freedom decoupled load cell (the third channel is supposed to be sensitive to the applied load component, F_3)

The readings per unit force values are easily yielded by the slopes of the linear interpolations of the experimental data points and the slopes of regression lines become the coefficients in the third column of the matrix **B**, i.e. $b_{13} \dots b_{63}$, Eq. 2.65. If the described procedure is repeated for all the load components, the entire matrix **B** can be obtained.

Upon calibration, when the cell is used to measure an unknown set of forces and moments, the outputs of the six channels can be processed by inverting the compliance matrix. If the strain gages are bonded in suitable locations and are adequately connected, the outputs of each channel are linearly independent from one another. After controlling the non-singularity of the compliance matrix **B** and its invertibility, the vector of the unknowns can be computed as in Eq. 2.66.

$$\Delta \mathbf{e} = \mathbf{B} \mathbf{f} \Leftrightarrow \mathbf{f} = \mathbf{B}^{-1} \Delta \mathbf{e} \Leftrightarrow \mathbf{f} = \mathbf{C} \Delta \mathbf{e} \tag{2.66}$$

The matrix **C**, the inverse of **B**, is commonly regarded as the *Calibration Matrix*. Otherwise, when strain gage location is badly designed, the risk is that the problem becomes ill-conditioned, which reflects in the singularity of **B** and in its determinant tending to zero.

Once the matrices **B** and **C** have been determined, it is possible to evaluate the response of the cell, in particular its decoupling, by computing suitable parameters, called *Coupling Coefficients* (C.C.). The value of the (ij)-th coefficient is yielded by Eq. 2.67.

$$C.C._{ij} = \frac{b_{ij} \cdot FS_j}{b_{ii} \cdot FS_i} \tag{2.67}$$

It can be regarded as the ratio between the output voltage of the i -th channel, when the j -th load component is acting (a component to which the i -th channel should be theoretically insensitive) and the output of the same channel, under the component it is designed to measure.

Values at full scale (FS_j, FS_i) must be considered in the application of Eq. 2.67. For a load cell with n degrees of freedom, $(n - 1)$ different $C.C.$ may be calculated for each channel. The grade of decoupling is indeed related to the values of the coupling coefficients, with low values indicating a good decoupling.

2.10.1 Load Cell for Three Forces

Measuring three forces along three perpendicular directions is a quite common requirement in robotic wrists. An accurate force measurement is required to prevent product damage or to make sure it is safely handled. The basic features of the cell in Fig. 2.35 may be summarized in its simple manufacturing process, in its sensitivity and in channel decoupling.

Regarding the first issue, a commercial annular tube (external diameter: 30 mm, internal diameter: 24 mm) was used. It is fully constrained at one end and connected to a small flat bar at the other one.

The forces, namely a longitudinal (F_x), a transverse (F_y) and a vertical load (F_z), are applied at its end. In order to increase the sensitivity of the three channels devoted to the measurement of the three forces, the cell is designed so that the applied strain gages are loaded under bending for every force component. If only the tube had been instrumented, there would have been a lack of sensitivity in the measurement of the axial component. For this reason a flat bar is connected and instrumented for the measurement of the axial component, converted into a bending load. The schemes of the bridges are in Fig. 2.36: channels A and C are half bridges and B is a full bridge.

The number of output channels equals the number of unknowns, thus the compliance and the calibration matrices are square (3×3). In order to achieve a good decoupling, each channel is devoted to the measurement of just one component and theoretically insensitive to all the others.

Channel A is sensitive to the longitudinal load, F_x : This force, up to 50 N causes the flat bar bending. Let H_b and S_b be the width and the thickness of the bar respectively: the bending resisting moment W_b ,¹⁴ holds:

$$W_b = \frac{1}{6} \cdot bh^2 \quad (2.68)$$

¹⁴ The resisting moment is the area moment of inertia I along the x -axis, over d , the distance from the neutral axis along y -axis on the section area in which the stress is calculated.

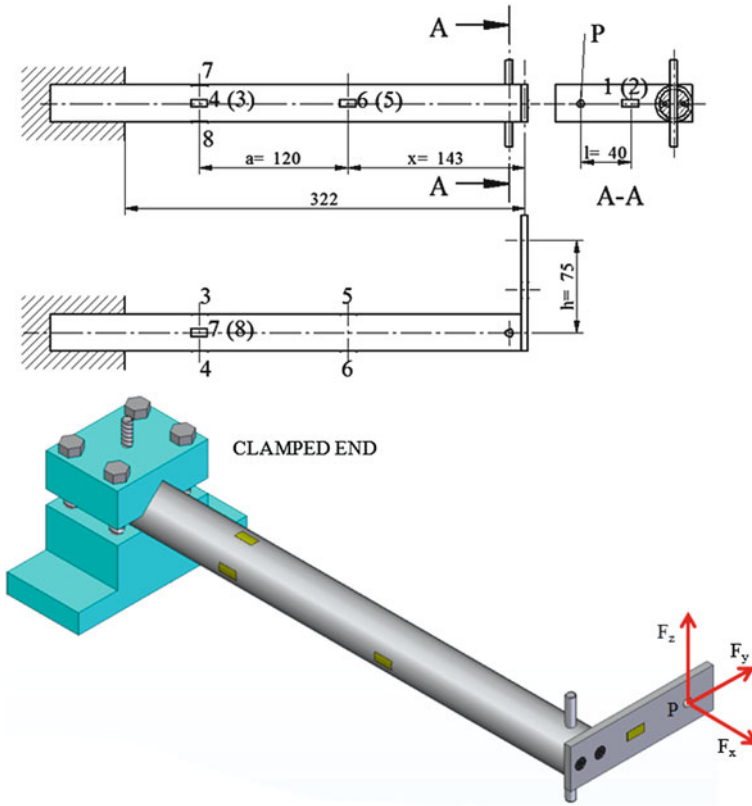


Fig. 2.35 Layout of a load cell for measuring three forces

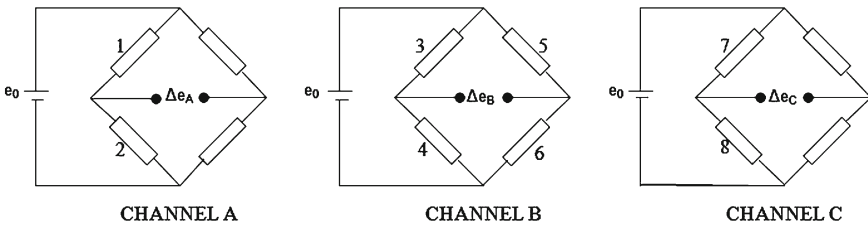


Fig. 2.36 Wheatstone bridge connections: channels A, B and C

The output voltage Δe_A on channel A is:

$$\Delta e_A = \frac{1}{4} \cdot K \cdot e_0 \cdot (\varepsilon_1 - \varepsilon_2) = \frac{1}{2} \cdot K \cdot e_0 \cdot \frac{F_x \cdot l}{EW_b} \quad (2.69)$$

Symbols K and e_0 in Eq. 2.69 stand for the gage factor and the supply voltage, whereas ε_1 and ε_2 indicate the strains under the strain gages 1 and 2 in Figs. 2.35 and 2.36. Finally, l indicates the moment distance in Fig. 2.35.

This output is theoretically independent of the other components: the transverse force F_y induces a normal stress on the flat bar, so that the strain reading is the same for the two strain gages and the resulting output is zero. The vertical force F_z implies bar bending: the strain gages 1 and 2 are in this case along the barycentric neutral axis: as a consequence, the measured strain is theoretically zero.

The force F_y , up to 200 N, is measured by the second half Wheatstone Bridge, indicated as B . Its output voltage, Δe_B , is computed:

$$\Delta e_B = \frac{1}{4} \cdot K \cdot e_0 \cdot (\varepsilon_3 - \varepsilon_4 + \varepsilon_6 - \varepsilon_5) = \frac{1}{2} \cdot K \cdot e_0 \cdot \frac{F_y \cdot a}{EW_b} \quad (2.70)$$

where a denotes the moment distance, as in Fig. 2.35. $\varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6$ stand for strains under the gages with numbers 3, 4, 5, 6 in the same figure.

The vertical load F_z generates a torsion moment transmitted to the tube. Torque is not detected by the previously mentioned strain gages, as they are aligned along the axial direction. The same load also has the effect of tube bending, however, the considered strain gages are applied on a barycentric plane: the detected strains are therefore zero. The longitudinal force F_x generates a constant bending moment on the tube: in this case strain gages with numbers 3 and 5 are under a non-zero tensile strain, whereas the other gages are under the same strain in compression. The output voltage is therefore zero.

The vertical load F_z , up to 200 N, is measured by the third channel C . As previously remarked, this force leads to a torsion of the tube, however this type of load is not detected by the axially aligned gages 7 and 8. The computation of the output voltage is shown in Equation:

$$\Delta e_C = \frac{1}{4} \cdot K \cdot e_0 \cdot (\varepsilon_7 - \varepsilon_8) = \frac{1}{2} \cdot K \cdot e_0 \cdot \frac{F_z \cdot (a + x)}{EW_b} \quad (2.71)$$

where x is a distance, Fig. 2.35, while ε_7 and ε_8 are the strains under the strain gages 7 and 8.

The axial load, F_x implies a normal load on the tube with identical strains detected by the gages with numbers 7 and 8, and therefore a negligible output. In addition, the tube bends, but the mentioned strain gages are bonded on a neutral plane and cannot detect any strain. Also the bending load, generated by the transverse force F_y , is not detected for the same reason.

The experimental tests were performed with a separate application of each load component. The outputs of the three channels were recorded simultaneously by a multi-channel acquisition device. The loads were applied by hanging calibrated masses. The tests were conducted with a step-by-step load increase up to maximum values, each trial was repeated for three times. The results are collected in diagrams

showing the output voltages of the three Wheatstone Bridges versus the applied loads (supply voltage: 2 V, gage factor: 2.09).

2.10.1.1 Loading by Axial Force F_x

The tube was oriented, so that the required load could be applied by calibrated masses under gravity. The results are shown in the diagram and in the histogram in Fig. 2.37. The highest output was retrieved for the first bridge A.

It is remarkable that the linear correlation coefficient (R^2) of the interpolating line related to bridge A is almost 1, indicating a perfect linearity. The R^2 coefficient is further from 1, considering the other channels, whose outputs (theoretically zero) are mainly due to experimental uncertainty. The histogram in the same Figure compares the experimental outputs to the theoretical predictions.

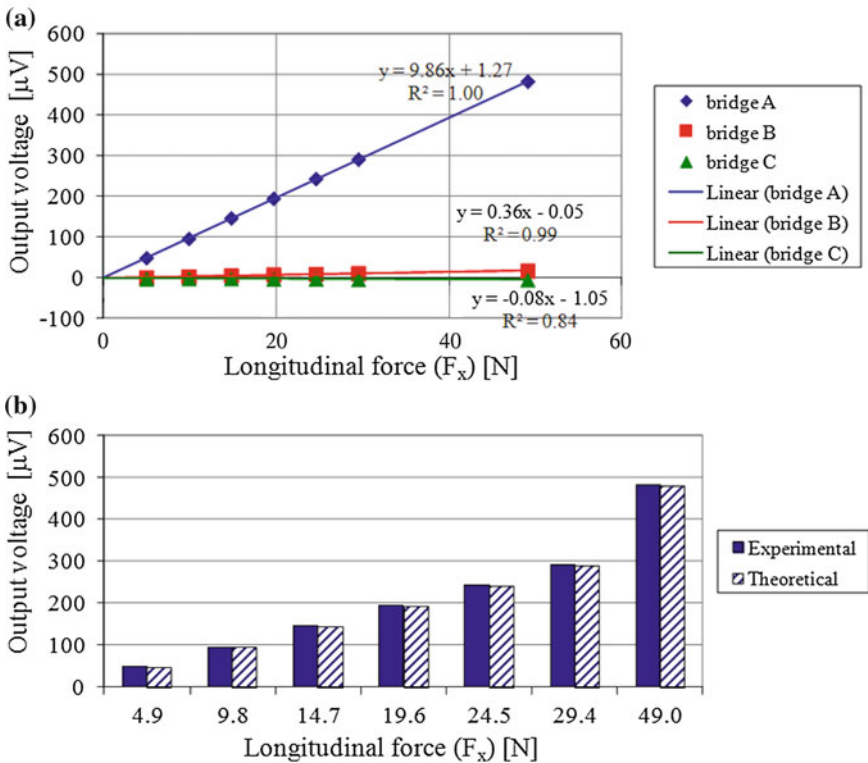


Fig. 2.37 a Output voltages on the three channels versus F_x entity; b comparison between the experimental outputs of channel A and theoretical predictions

2.10.1.2 Loading by the Transverse Force F_y

For the application of the transverse load the beam was properly oriented, and the experimental results are shown in Fig. 2.38.

The first diagram at the top indicates the most significant output as being that of the second channel (*B*). However, the output of the first channel (*A*) is also remarkable (up to $72 \mu\text{V}$) and much higher than expected. According to the theoretical model, under the hypothesis of negligible displacements and rotations, the strain gages with numbers 1 and 2 should be loaded by a normal load. Moreover, as they are mounted on adjacent arms of the bridge, the overall output should be zero. It can be observed that the vertical displacement of the tube is not completely negligible (in the order of mm): consequently, the flat plate experiences a bending deformation.

The strains due to bending are therefore retrieved by the strain gages and in addition, the output signal is doubled, due to the half-bridge connection. This can be regarded as a secondary effect, which can be anyway cancelled by the cell calibration, with a consequent determination of the calibration matrix. In this case, also the

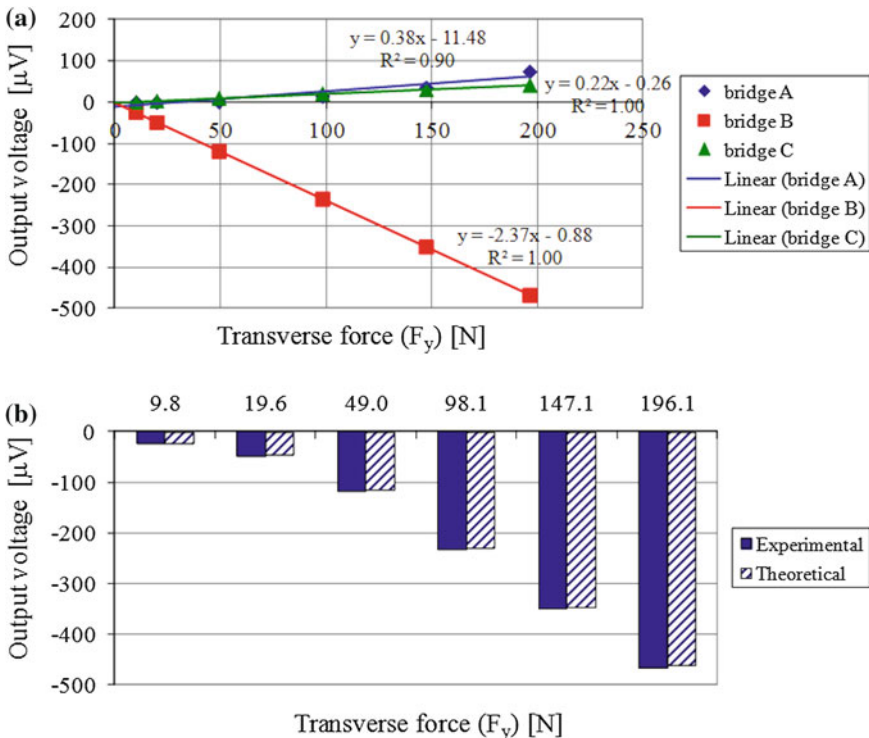


Fig. 2.38 **a** Output voltages on the three channels versus F_y entity; **b** comparison between the experimental outputs of channel *B* and theoretical predictions

regression lines with reference to the most involved channel (*B*) have a high linear correlation coefficient R^2 (close to one).

Experimental and theoretical results are also compared in the histogram in Fig. 2.38 in the bottom, with errors below 2%.

2.10.1.3 Loading by the Vertical Force F_z

The vertical load was applied under another orientation of the cell. The results, Fig. 2.39a, show a good decoupling and a very good linearity in the response of the third bridge channel *C*. The data are compared to the theoretical outcomes in the histogram in Fig. 2.39b with a very good agreement (errors in the order of 2%).

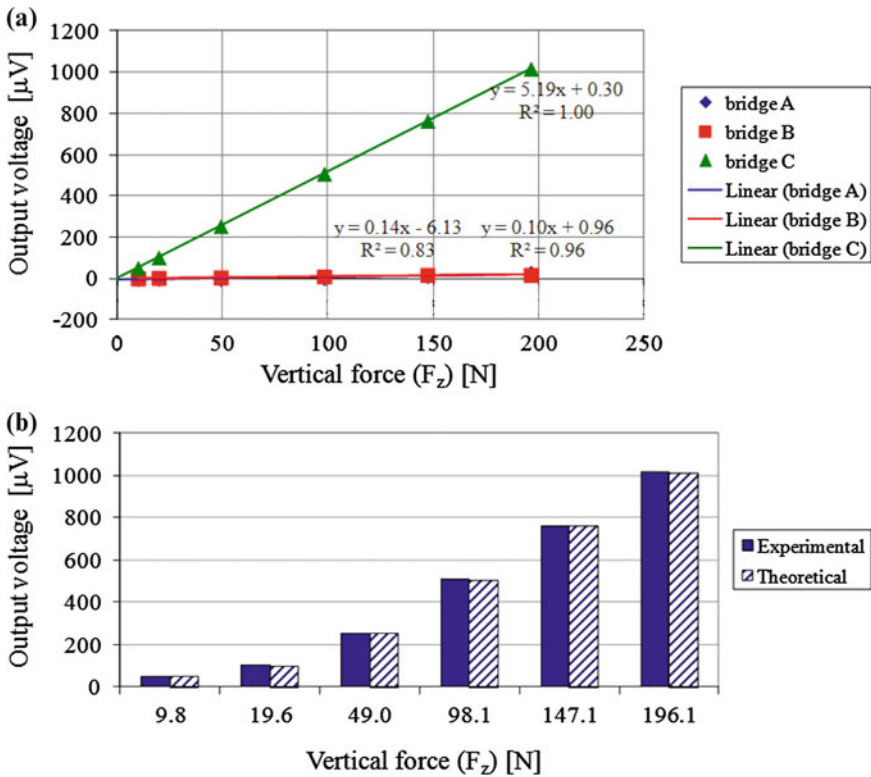


Fig. 2.39 a Output voltages on the three channels versus F_z entity. b Comparison between the experimental outputs of channel *C* and theoretical predictions

Concluding, it can be pointed out that:

- The calibrated cell generally has a good decoupling, with an exception for channel A, designed to measure F_x , being slightly sensitive to F_y .
- The experimental results are linearly distributed with very high values of the linear correlation coefficients (R^2).
- The outputs of the three channels increase with the entity of the applied load with negligible offsets.

The results of the calibration tests were processed, considering the slopes of the linear distributions (expressed in $\mu\text{V}/\text{N}$) mentioned above. These coefficients, having the meaning of output voltage per unit applied load, can be regarded as the coefficients of the compliance matrix \mathbf{B} .

Exercise 2.3 (*Determination of the Compliance and Calibration matrices*) The slopes determined under the first load component F_x were regarded as the coefficients of the first column of the compliance matrix (\mathbf{B}). In the same way, the slopes under the second F_y and the third F_z components were incorporated into the second and the third columns of the same matrix.

The relationship between the output voltages and the load array, along with the matrix (\mathbf{B}) is reported in the following Eq. 2.72.¹⁵

$$\Delta \mathbf{e}_{mV} = \mathbf{B} \mathbf{f}_N \quad (2.72)$$

$$\begin{pmatrix} \Delta e_A \\ \Delta e_B \\ \Delta e_C \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \cdot \begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix}$$

Numerically:

$$\begin{pmatrix} \Delta e_A \\ \Delta e_B \\ \Delta e_C \end{pmatrix} = \begin{pmatrix} 9.9 \times 10^{-3} & 3.8 \times 10^{-4} & 1.4 \times 10^{-4} \\ 3.6 \times 10^{-4} & -2.4 \times 10^{-3} & 9.6 \times 10^{-5} \\ -7.0 \times 10^{-5} & 2.2 \times 10^{-4} & 5.2 \times 10^{-3} \end{pmatrix} \cdot \begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix}$$

The matrix \mathbf{B} has a non-zero determinant and is therefore invertible. Its inverse, matrix \mathbf{C} , is determined:

$$\mathbf{f}_N = \mathbf{C} \Delta \mathbf{e}_{mV} \quad (2.73)$$

$$\begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \cdot \begin{pmatrix} \Delta e_A \\ \Delta e_B \\ \Delta e_C \end{pmatrix}$$

¹⁵ In order to avoid too low coefficients of the Calibration matrix, it is common use to express the output voltages in mV. Therefore, the previously determined slopes are divided by 10^3 .

$$\begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix} = \begin{pmatrix} 100.8 & 15.8 & -3.1 \\ 15.3 & -418.0 & 7.3 \\ 0.8 & 17.8 & 192.5 \end{pmatrix} \cdot \begin{pmatrix} \Delta e_A \\ \Delta e_B \\ \Delta e_C \end{pmatrix}$$

The calibration matrix \mathbf{C} is very close to the condition for a diagonal matrix. In fact, the terms on the diagonal are higher than the other by about one order of magnitude. The level of decoupling of the design load cell is further evaluated by the determination of the coupling coefficients, as in Eq. 2.67 at page 73. The maximum values at full scale are: $FS_{F_x} = 49 \text{ N}$, $FS_{F_y} = 196 \text{ N}$, $FS_{F_z} = 196 \text{ N}$. The computed coefficients are shown in Eq. 2.74.

$$\begin{aligned} C.C.s &= \begin{pmatrix} \text{---} & \frac{b_{12} \cdot FS_{F_y}}{b_{11} \cdot fS_{F_x}} & \frac{b_{13} \cdot fS_{F_z}}{b_{11} \cdot fS_{F_x}} \\ \frac{b_{21} \cdot FS_{F_x}}{b_{22} \cdot FS_{F_y}} & \text{---} & \frac{b_{23} \cdot FS_{F_z}}{b_{22} \cdot FS_{F_y}} \\ \frac{b_{31} \cdot FS_{F_x}}{b_{33} \cdot FS_{F_z}} & \frac{b_{32} \cdot FS_{F_y}}{b_{33} \cdot FS_{F_z}} & \text{---} \end{pmatrix} \\ &= \begin{pmatrix} \text{---} & 15.38 \% & 5.83 \% \\ -3.79 \% & \text{---} & -4.05 \% \\ -0.36 \% & 4.21 \% & \text{---} \end{pmatrix} \end{aligned} \quad (2.74)$$

The values are all within a few percentage points, except for the highest one, $C.C._{12} = 15.38 \%$. This shows that the first channel is slightly dependent on transverse loads, due to the secondary effects.

2.10.2 Load Cell for Two Forces and One Torsional Moment

The design of this load cell is inspired by the same principles discussed above. The aim was to design a decoupled cell with good accuracy and sensitivity and that could be easily calibrated and tested.

The same tube as the load cell of Sect. 2.10.1 is clamped at one end and loaded at the other. Loads are a transverse and a vertical force plus a torsional moment.

This type of cell may also have applications in the field of robotics, to determine force on gripping devices and to detect disturbing torsional components. All the strain gages measure strains due to bending generated by the mentioned loads. For measuring torsion, four strain gages were bonded along $+45^\circ$ and -45° helical lines and connected to the four arms of a full Wheatstone bridge.

The cell was conceived as theoretically decoupled: each bridge channel is sensitive to a single component and insensitive to the others, Figs. 2.40 and 2.41. The application point of forces F_y and F_z and of torsional moment M_x is indicated by Q .

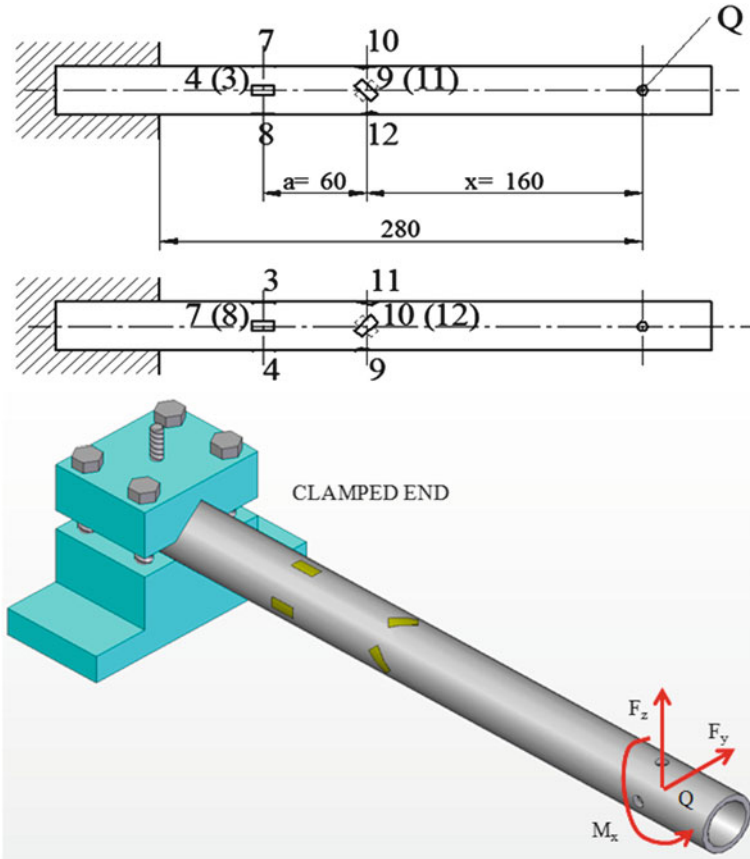


Fig. 2.40 Load cell for measuring two forces and a torsional moment: layout, main dimensions, instrumentation

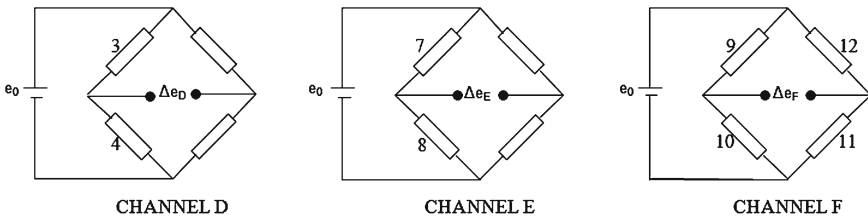


Fig. 2.41 Strain gage Wheatstone bridge connections: channels *D*, *E* and *F*

The first channel, corresponding to the bridge arm *D*, is sensitive to the transverse force F_y which bends the tube. The output voltage can be computed by Eq. 2.75, where K is the gage factor, e_0 is the supply voltage, and ε_3 and ε_4 respectively stand for the strains retrieved by the gages with numbers 3 and 4. The distances a and x

are defined in Fig. 2.40, whereas E and W_b are the material Young's modulus and the cross section bending resisting moment.

$$\Delta e_D = \frac{1}{4} \cdot K \cdot e_0 \cdot (\varepsilon_3 - \varepsilon_4) = \frac{1}{2} \cdot K \cdot e_0 \cdot \frac{F_y \cdot (a + x)}{EW_b} \quad (2.75)$$

The vertical load F_z is not detected by this channel, as the strain gages 3 and 4 are on the neutral plane for bending. Torsion is definitely ineffective for strain gages with longitudinal orientation. The same happens for the fourth channel, bridge arm E , designed to measure the vertical force F_z . The computation of the output voltage is in Eq. 2.76, where ε_7 and ε_8 respectively stand for the strains recorded by the gages with numbers 7 and 8.

$$\Delta e_E = \frac{1}{4} \cdot K \cdot e_0 \cdot (\varepsilon_7 - \varepsilon_8) = \frac{1}{2} \cdot K \cdot e_0 \cdot \frac{F_z \cdot (a + x)}{EW_b} \quad (2.76)$$

The last channel, bridge arm F , with four active gages on $+45^\circ$ and -45° helical lines, is sensitive to torsion M_x . The moment is applied by hanging a mass (up to 196 N) at a distance of 75 mm. The output voltage is computed by Eq. 2.77, where ε_9 , ε_{10} , ε_{11} and ε_{12} respectively stand for the strains recorded by the gages with numbers 9, 10, 11 and 12.

$$\Delta e_F = \frac{1}{4} \cdot K \cdot e_0 \cdot (\varepsilon_9 - \varepsilon_{10} + \varepsilon_{11} - \varepsilon_{12}) = K \cdot e_0 \cdot \frac{1 + \nu}{E} \cdot \frac{M_x}{W_t} \quad (2.77)$$

This output, where ν indicates the material Poisson's coefficient and W_t the torsional resisting moment, is insensitive to the bending moments generated by the forces F_y and F_z .

2.10.2.1 Loading by the Transverse Force F_y

Calibration results are shown in Fig. 2.42. The output of the first channel D is much higher than the others. The regression lines have correlation coefficient R^2 very close to one. The histogram of comparison between the experimental yields and the numerical predictions shows very good agreement.

2.10.2.2 Loading by the Vertical Force F_z

The results obtained upon the application of the vertical load are shown in Fig. 2.43. The output voltage of the second bridge, (E), is now the most significant with respect to the other ones. There is very good linearity and an excellent agreement between experimental and theoretical results, with errors in the order of just 1%.

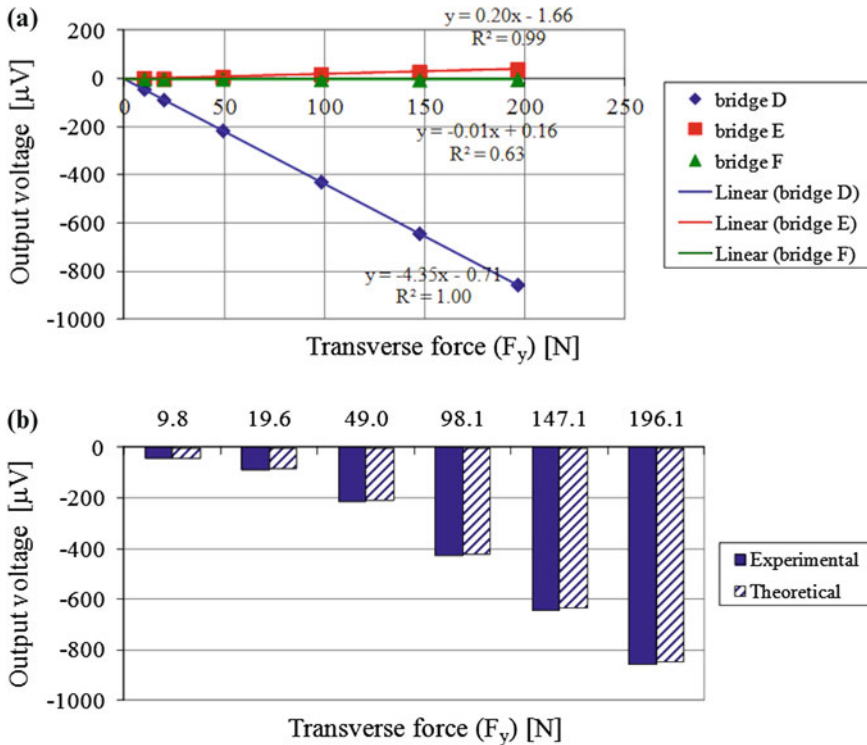


Fig. 2.42 a Output voltages on the three channels versus F_y entity; b comparison between the experimental outputs of channel D and theoretical predictions

2.10.2.3 Loading by the Torsional Moment M_x

The application of a pure torsional moment requires a specific test rig. Some possible solutions are sketched in Fig. 2.44. A single vertical force is applied by gravity and the outputs of the three channels were recorded. The force was applied, so that the longitudinal distance from the clamping device was the same as in the F_z force application trial (described in Sect. 2.10.2.2), so that the experimental results could be made dependent on the same bending moment as in the previous testing.

Based on the *Principle of the Superposition of the Effects*, the contribution of bending moment could be cancelled by computing the difference between the current outputs and those reported in Sect. 2.10.2.2.

The test results are shown in Fig. 2.45: they account for both torsional and bending moments, therefore the outputs are significant on channels E and F. A double scale is reported along the horizontal axis, thus indicating both the value of the force (F_z) and that of the induced torsional moment (M_x). The application of the force is far away from the longitudinal axis, at a distance of 75 mm.

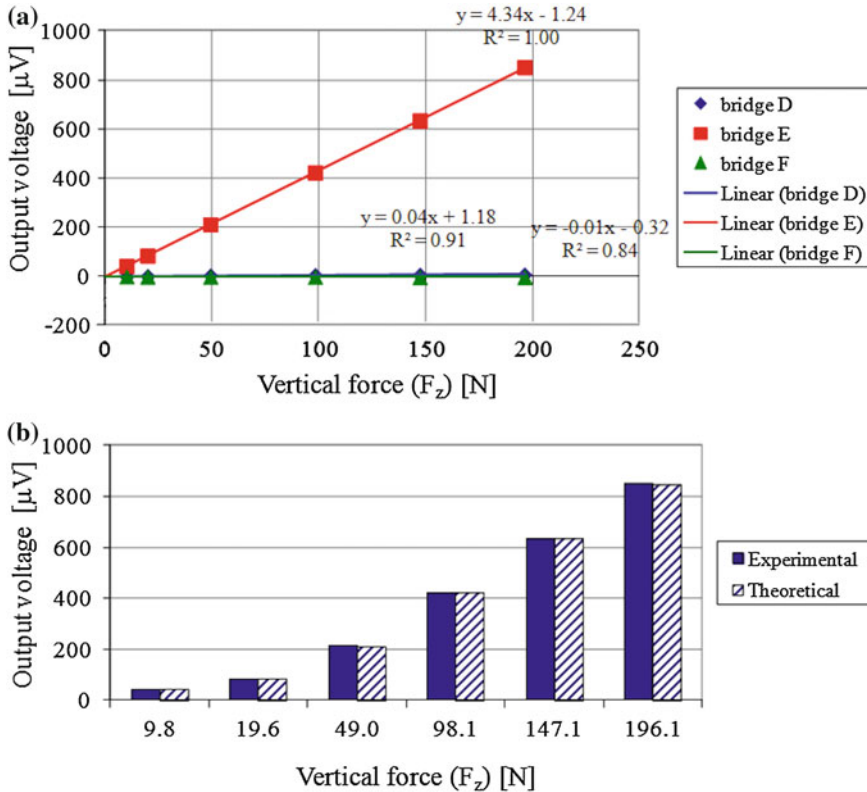


Fig. 2.43 a Output voltages on the three channels versus F_z entity; b Comparison between the experimental outputs of channel E and theoretical predictions

The described procedure of detracting the outputs depending on the bending moment, from those depending on both bending and torsional moments led to the results depicted in Fig. 2.46. These results are now dependent on (pure) torsion only: it can be remarked that this type of load is detected only by the third channel (F). The values are consistent with theoretical predictions, apart from a small error (about 3%). It can be pointed out that:

- The calibrated cell generally has a good decoupling.
- The experimental results are linearly distributed with very high values of the linear correlation coefficients (R^2).
- The outputs of the three channels increase with the entity of the applied load with negligible offsets.

Results of the calibration tests are processed, considering the slopes of the linear distributions (expressed in $\mu V/N$) mentioned above. These coefficients, having the meaning of output voltage per unit of applied load, can be regarded as the coefficients of the compliance matrix **B**.

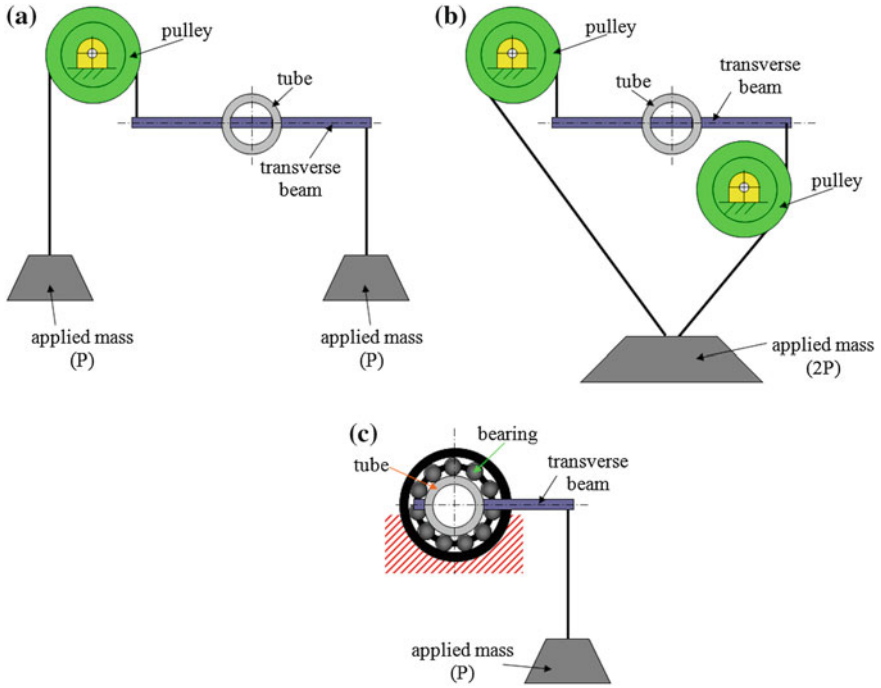


Fig. 2.44 Several solutions for pure torsional moment application: **a, b** using return pulleys; **c** using a radial bearing

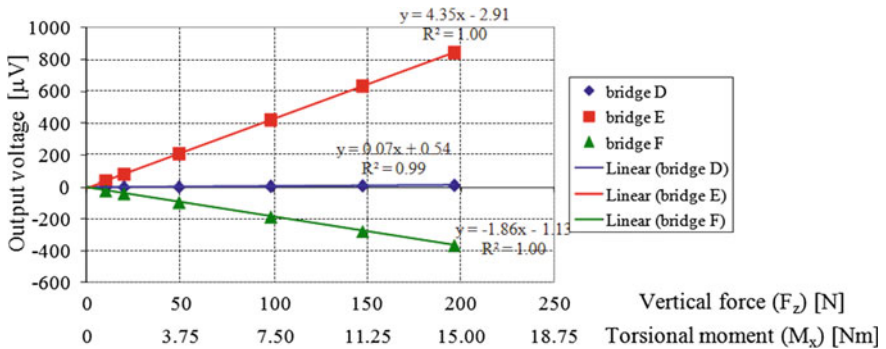


Fig. 2.45 Output voltages on the three channels versus F_z and M_x

Exercise 2.4 (*Determination of the Compliance and Calibration matrices*) The same procedure was followed in the processing of the experimental data and in the determination of the compliance matrix and related coefficients. In particular, the slopes of the regression lines obtained under the first load component (F_y) enter the first matrix column. In the same way, the slopes of the regression lines

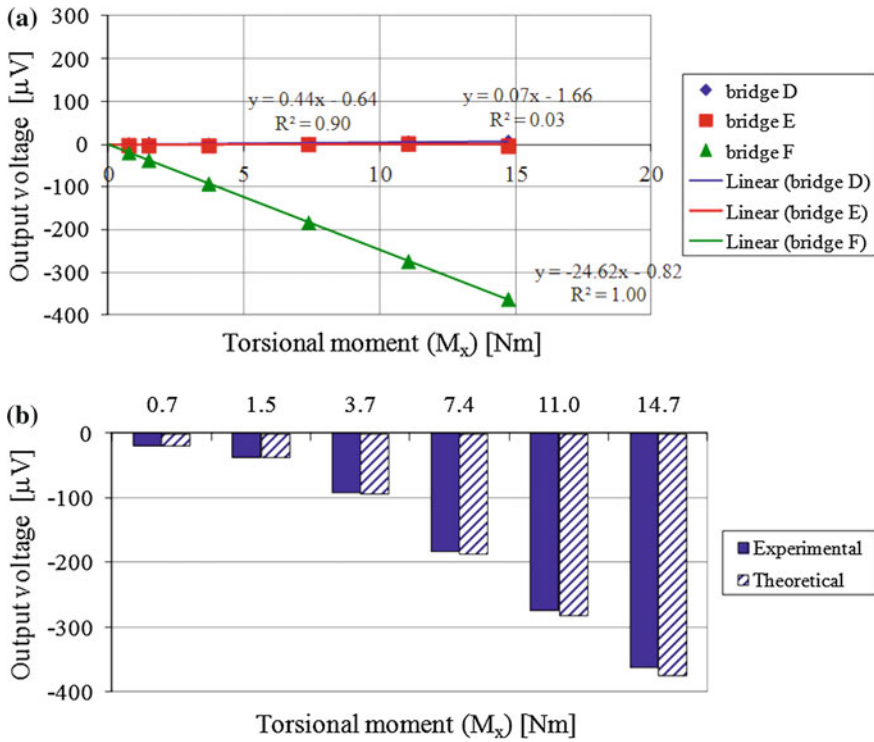


Fig. 2.46 **a** Output voltages on the three channels versus M_x entity; **b** comparison between the experimental outputs of channel F and theoretical predictions

determined under F_z and M_x were incorporated into the second and the third columns respectively.

The relationship between the output voltages and the load array is reported in Equation:

$$\Delta \mathbf{e}_{mV} = \mathbf{B} \mathbf{f}_N \tag{2.78}$$

$$\begin{pmatrix} \Delta e_D \\ \Delta e_E \\ \Delta e_F \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \cdot \begin{pmatrix} F_y \\ F_z \\ M_x \end{pmatrix}$$

$$\begin{pmatrix} \Delta e_D \\ \Delta e_E \\ \Delta e_F \end{pmatrix} = \begin{pmatrix} -4.4 \times 10^{-3} & 3.5 \times 10^{-5} & 4.4 \times 10^{-4} \\ 2.0 \times 10^{-4} & 4.3 \times 10^{-3} & 6.9 \times 10^{-5} \\ -1.3 \times 10^{-5} & -1.1 \times 10^{-5} & -2.5 \times 10^{-2} \end{pmatrix} \cdot \begin{pmatrix} F_y \\ F_z \\ M_x \end{pmatrix}$$

The matrix \mathbf{B} has full rank and is therefore invertible. Its inverse, matrix \mathbf{C} , is reported in Eq. 2.79. The output voltages are turned into mV

$$\mathbf{f}_N = \mathbf{C} \Delta \mathbf{e}_m \mathbf{V} \tag{2.79}$$

$$\begin{pmatrix} F_y \\ F_z \\ M_x \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \cdot \begin{pmatrix} \Delta e_D \\ \Delta e_E \\ \Delta e_F \end{pmatrix}$$

$$\begin{pmatrix} F_y \\ F_z \\ M_x \end{pmatrix} = \begin{pmatrix} -229.6 & 1.9 & -9.41 \\ 10.8 & 230.3 & 0.8 \\ 0.1 & -0.1 & -40.6 \end{pmatrix} \cdot \begin{pmatrix} \Delta e_D \\ \Delta e_E \\ \Delta e_F \end{pmatrix}$$

The calibration matrix \mathbf{C} is almost diagonal: the terms on the main diagonal are much greater (by about two orders of magnitude) than the other ones. The good level of decoupling is well confirmed also by the computation of the coupling coefficients, as in Eq. 2.67 at p. 72. The following values at full scale are considered for calculation 196 N for F_y and F_z and 14.70 Nm for M_x . The computed coefficients are shown in Eq. 2.80.

$$\begin{aligned} C.C.s &= \begin{pmatrix} \text{---} & \frac{b_{12} \cdot FS_{F_z}}{b_{11} \cdot FS_{F_y}} & \frac{b_{13} \cdot FS_{M_x}}{b_{11} \cdot FS_{F_y}} \\ \frac{b_{21} \cdot FS_{F_y}}{b_{22} \cdot FS_{F_z}} & \text{---} & \frac{b_{23} \cdot FS_{M_x}}{b_{22} \cdot FS_{F_z}} \\ \frac{b_{31} \cdot FS_{F_y}}{b_{33} \cdot FS_{M_x}} & \frac{b_{32} \cdot FS_{F_z}}{b_{33} \cdot FS_{M_x}} & \text{---} \end{pmatrix} \\ &= \begin{pmatrix} \text{---} & -0.81 \% & -7.40 \% \\ 4.69 \% & \text{---} & 1.16 \% \\ 0.07 \% & 0.06 \% & \text{---} \end{pmatrix} \end{aligned} \tag{2.80}$$

The values are all within few percentage points. The hugest one, $|C.C.13| = 7.40 \%$ indicates that the first channel (D) has a slight sensitivity to the torsional moment M_x . It may be due to a positioning or more likely an alignment error of the strain gages with numbers 7 and 8, which should theoretically be longitudinally oriented.

2.10.3 Load Cells at Six Degrees of Freedom

Two alternatives case studies of cells were presented to identify six load and moment components in a piece of sports equipment (actions transmitted by the ski boot to the ski), Fig. 2.47a.

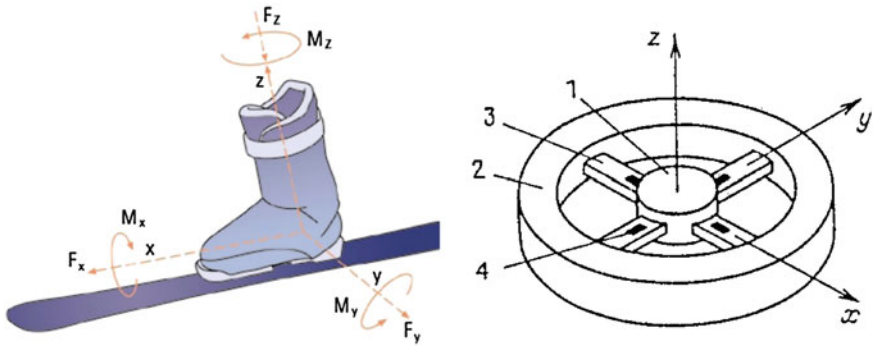


Fig. 2.47 The six actions transmitted by a ski boot to the ski and comparison with a classic scheme of a generic cell with more degrees of freedom developed for a robot hand

2.10.3.1 Statically Determinate Cell

The cell consists of seven bending micro-cells, each of which is calibrated for a force measurement, Fig. 2.48 through cantilever small beams. The cell is divided in sub-cells: a frontal sub-cell and a rear sub-cell. A central cell completes the device [48, 49].

The first frontal sub-cell measures the component f_x in the direction of motion along the longitudinal axis x of the ski, the two parallel micro-cells measure two vertical forces in the vertical direction z which, together with the homologous sub-cells in the rear part, allow the independent measurement of the force f_z as algebraic sum of the four values, as well as the measurement of the *pitching moment* (with the moment-axis in the transverse direction y) and the *rolling moment* (with moment-axis along the longitudinal x axis).

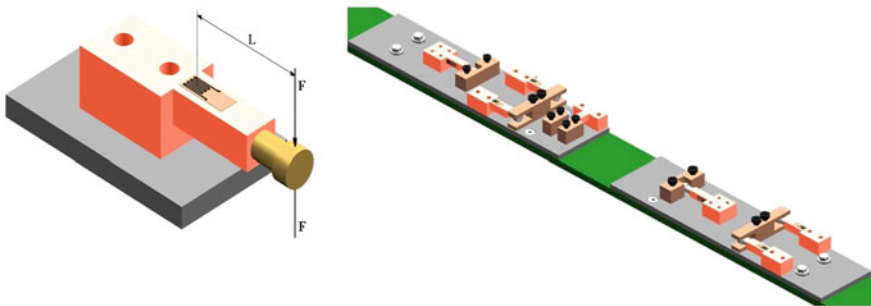


Fig. 2.48 Scheme of the isostatic cell with six degrees of freedom with seven micro-cells in bending for the recording of six signals [29, 50]

The central micro-cell, together with the homologous in the back part, allows the determination of the lateral force along the y axis and also the torque moment (with moment-axis along the vertical axis z).

The calibration of the cell needs the application of six independent actions (three forces and three moments). They make it possible to determine a matrix of compliance, also called *calibration matrix*, which connects the output signals from the measuring channels to the six actions. In this case a square matrix is obtained that must be inverted, for identifying the unknowns actions from the measurements. The inverse operation is mathematically possible, without any difficulty.

2.10.3.2 Statically Indeterminate Cell

Conversely, a second solution, designed for safety reasons with an indeterminate structure with six unknowns and eight measurements is shown in Fig. 2.49.

The cell has, Fig. 2.50:

- 6 Actions to evaluate (3 Forces and 3 Moments).
- 24 Sensors.
- 8 Wheatstone bridges (4 half- and 4 full bridges).
- 8 Equations with 6 unknowns. (overdetermined problem).

The compliance matrix obtained from the calibration is no longer square due to an excess of experimental data with respect to the unknowns. This condition is essential for finding acceptable solutions for ill-conditioned problems and is effective for making the measuring instrument sufficiently accurate, precise and robust (insensitive to noise).

Exercise 2.5 (*Calibration of a 6-degrees of freedom load cell*) The calibration was done by the application of known forces and moments to the cell, as in the example shown in Fig. 2.51, in order to determine first the compliance matrix, Eq. 2.81.

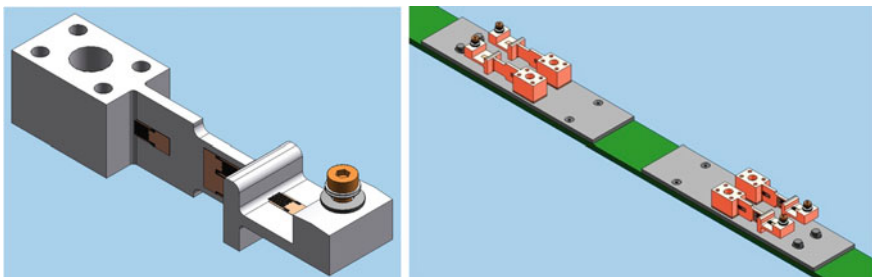


Fig. 2.49 Scheme of a hyperstatic cell at eight degrees of freedom for recording of six signals. Overdetermined problem

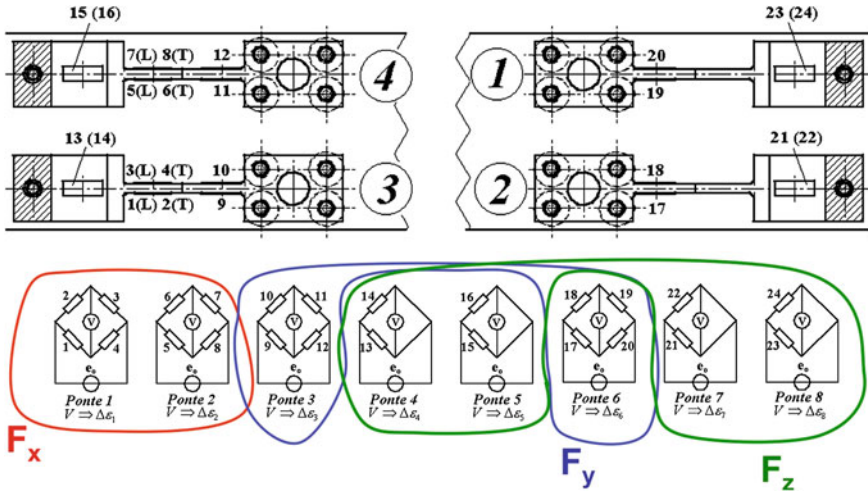


Fig. 2.50 Strain gage location on the four micro-cells and circuit diagram with eight degrees of freedom for the recording of six actions by overdetermined data [29, 51]



Fig. 2.51 Calibration for the cell mounted on a ski

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & \dots & \dots & \dots & b_{16} \\ b_{21} & b_{22} & \dots & \dots & \dots & b_{26} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ b_{81} & b_{82} & \dots & \dots & \dots & b_{86} \end{pmatrix} \cdot \begin{pmatrix} F_x \\ F_y \\ F_z \\ M_x \\ M_y \\ M_z \end{pmatrix} \quad (2.81)$$

According to Sect. 2.10, the relationship between the output and the input being the following:

$$\varepsilon = \mathbf{B} \cdot \mathbf{F}$$

the pseudo-inversion of the rectangular matrix gives:

$$\mathbf{f} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \boldsymbol{\varepsilon} = \mathbf{B}^{-g} \boldsymbol{\varepsilon}$$

see Eq. 1.23 in Chap. 1.

The following matrix \mathbf{B}^{-g} is the *Calibration Matrix* that is shown in Eq. 2.82.

$$\begin{pmatrix} F_x \\ F_y \\ F_z \\ M_x \\ M_y \\ M_z \end{pmatrix} = \begin{pmatrix} 2.35 & 0.74 & 0.18 & 0.95 & -0.75 & -0.12 & -0.70 & 0.75 \\ 0.05 & 0.02 & -0.05 & 0.02 & -0.02 & -0.05 & -0.02 & 0.016 \\ 0.28 & 0.09 & 0.03 & -0.03 & -0.26 & -0.02 & -0.23 & -0.06 \\ 2.006 & 0.002 & -0.001 & -0.002 & 0.003 & -0.002 & -0.007 & 0.007 \\ 0.27 & 0.09 & 0.02 & 0.15 & -0.05 & -0.01 & -0.12 & 0.05 \\ 0 & -10^{-4} & 0.01 & -0.002 & 9 \times 10^{-4} & -0.01 & 0.002 & -2 \times 10^{-4} \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \end{pmatrix} \quad (2.82)$$

After the calibration matrix was calculated, the load cell was verified, comparing the assigned values of the three forces and the three moments with the six corresponding values *identified* through the measurements.

Comparison were performed as shown in diagrams of Fig. 2.52. In this case, discrepancies are present, especially on the M_y moment (error around 11 %) and on the F_x force (error up to 100 %).

Further tests were performed to investigate the repeatability of the results in dynamic application.

The cell weaknesses can be thus summarized:

- Lack of accuracy, mainly in the measurement of the longitudinal force F_x , and of the M_y moment.
- Lack of repeatability and signal conditioning, thus of precision, since small uncertainties in the estimation of the strains lead to much greater errors in the identification of forces and moments.

This case is a typical example of an overdetermined (the number of channels exceeds that of the measured components) ill-posed problem.

The next exercise tackles the questions posed above and shows possible strategies to improve the calibration of this kind of sensors in order to improve their accuracy and robustness.

An important improvement can derive from the regularization methods [52–54] of ill-posed problems, such as the *Tikhonov-Miller*, *Tikhonov-Phillips* algorithms. The first one is also known as the *Regularized Least Squares (RLS)* method. A further option consists in the *Singular Value Decomposition (SVD)*. Further information can be found in Chap. 1.

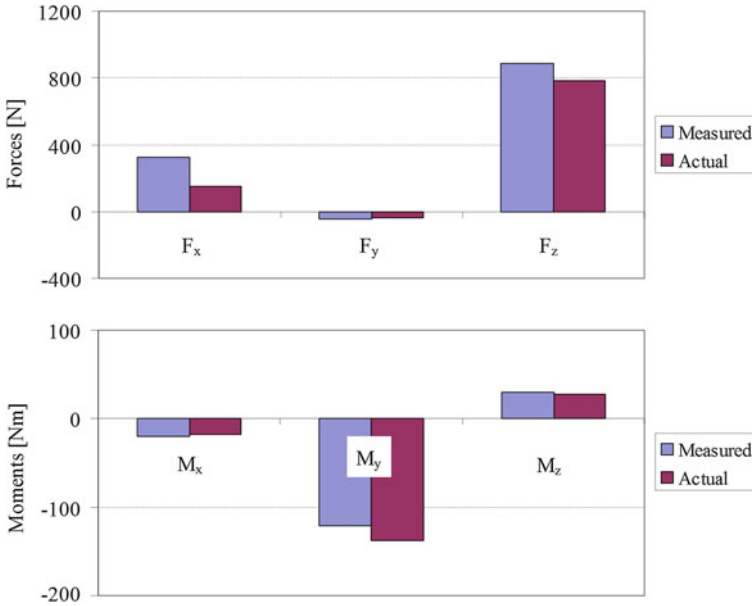


Fig. 2.52 Comparison between actual and measured force and moments before regularization

2.10.4 Regularization of the Six-Degrees Load Cell

The results of the previous calibration tests were first analyzed to determine the possible cause for the errors introduced in force estimation.

2.10.4.1 Lack of Accuracy of the Cell

The first issue to be considered was the lack of accuracy of the cell. The results of the calibration tests were analyzed to determine the possible cause for the errors introduced in force estimation.

The load cell was designed in order to achieve a good decoupling on the six measured components. This goal was accomplished by a suitable design of the sensing element, a suitable location of strain gages and a proper Wheatstone bridge connection. As stated before (see Sect. 2.10), this is a usually recommended design trend, to make sure of full sensitivity to all the acting forces and moments.

However, the results of the calibration tests denoted some unexpected trends, especially considering the channel which theoretically should have had a zero reading.

In these cases the output was obviously not exactly zero, but the output-load relationship was often noisy and quite difficult to linearly interpolate.

For instance, the responses of channels 1 and 2 versus the applied bending moment M_y are sketched in Fig. 2.53. These two channels are designed to measure the F_x component and theoretically insensitive to M_y . The strain readings are all repeatable and not particularly high (lower than $15 \mu\epsilon$).

However, the followed trend seems to be dependent on the sign of the applied moment, probably due to mechanical hysteresis in the load transmission to the sensing elements. The graphs in Fig. 2.54 emphasize another strain-load trend which is not suitable for linearization.

This figure refers to the responses of channels 4, 5, 7, 8, sensitive to M_y and to transverse force F_y . The trend is indeed linear, however a not negligible offset (about $45 \mu\epsilon$) is introduced. Not taking these offsets into account can be the primary cause of cell insufficient accuracy.

As a consequence of this first analysis, the experimental data were reprocessed without neglecting all the offset terms encountered in data linearization.

The formulas in Eq. 2.83 show how the offset term may be incorporated into the equation that gives the force vector \mathbf{f} as a function of the current strain readings over the eight channels ϵ .

The total offset is indicated as ϵ_0 , while superscripts are added to indicate the offset terms related to the six force and moment components.

$$\epsilon = \mathbf{B} \cdot \mathbf{f} + \epsilon_{0_{F_x}} + \epsilon_{0_{F_y}} + \epsilon_{0_{F_z}} + \epsilon_{0_{M_x}} + \epsilon_{0_{M_y}} + \epsilon_{0_{M_z}} = \mathbf{B} \cdot \mathbf{f} + \epsilon_0 \Leftrightarrow (2.83)$$

$$\Leftrightarrow \epsilon - \epsilon_0 = \mathbf{B} \cdot \mathbf{f} \Leftrightarrow$$

$$\Leftrightarrow \mathbf{f} = \mathbf{C} \cdot (\epsilon - \epsilon_0) = \mathbf{C} \cdot \epsilon'$$

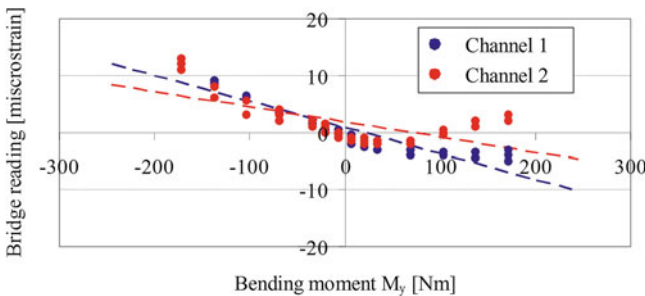


Fig. 2.53 Noise affecting signal with theoretical zero reading

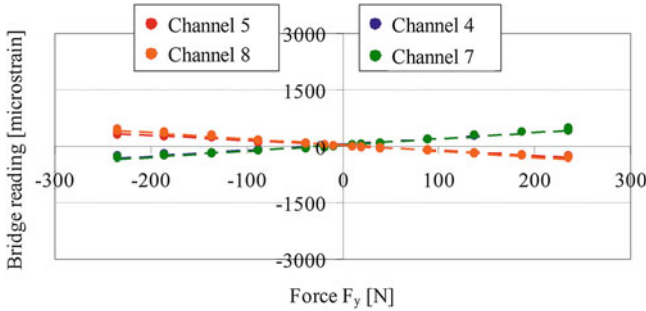


Fig. 2.54 A further example of noise affecting signal with theoretical zero reading

2.10.4.2 Lack of Robustness and Precision

The second issue regarded the lack of robustness and precision in load estimation. A response, where a huge error in the estimation of mechanical actions arises from even small errors in strain measurement, is typically regarded as ill-conditioning. In some references, e.g. in [52], it is remarked that an overdetermined problem, where the number of channels or measuring points is higher than the number of loads, is usually ill-conditioned. An acceptable solution is usually provided [55, 56] by the *Tikhonov-Phillips* [52–54] regularization algorithm. It can be regarded as a combination of the minimum-length with the least-squares solution, minimizing the object function, and is going to be applied in the following. As mentioned, alternative techniques consist in the application of the *Tikhonov-Miller* algorithm or of the *Singular Value Decomposition*. However, the first one has the drawback of requiring a priori information about the solution. On the other hand, the application of the second one must be usually integrated in the solution of the inverse problem and has the outcome of a lack of input reproducibility.

$$J = \|\varepsilon - \mathbf{B} \cdot \mathbf{f}\|^2 + \alpha \|\mathbf{f}\|^2 = \min \tag{2.84}$$

The solution of the least-squares problem in Eq. 2.84 gives the vector of the applied forces and moments, \mathbf{f} , as in Eq. 2.85. The symbol \mathbf{I} stands for the Identity matrix.

$$\mathbf{f} = (\mathbf{B}^T \mathbf{B} + \alpha \mathbf{I})^{-1} \mathbf{B}^T \cdot \varepsilon' \tag{2.85}$$

The relationships in Eqs. 2.84 and 2.85 contain a further term, α , commonly indicated as *damping factor*. The key issue in the application of the *Tikhonov-Phillips* regularization algorithm, consists in the determination of a suitable value for α . In Refs. [52–54, 57] it is remarked that this is usually between 0 and 1 and that it must be high enough to prevent ill-conditioning. In particular, α must have the smallest value to avoid ill-conditioning, a too high value is also unacceptable, as it would have a detrimental effect on solution accuracy. Its determination can be tackled by

two general purpose methods: the *L-Curve* and the *Generalized Cross Validation (GCV)* [58–62]. However, these two criteria, despite their good efficiency, are not suitable to all problems. An alternative approach that appears to be more suitable for the calibration and regularization of overdetermined load cells is suggested in the following section [63].

2.10.5 Determination of the Damping Factor

In order to determine the suitable value for α , two numerical tools can be developed, to simulate experimental uncertainty and to globally estimate error in the simultaneous measurement of more variables. The general goal is to find the lowest value of α , so that the results are sufficiently accurate, despite a noisy strain measurement.

For the sake of clarity, the method is applied with reference to the examined case study of three forces and three moments, but it can be used in general cases.

1. The first step was to consider some sample readings obtained in lab under combinations of loads. In order to randomly generate noisy signals, normal distributions were applied to the eight readings. The standard deviation was estimated as 10 % of the measured value, as this percentage can be regarded as a common estimation of experimental uncertainty in strain gage measurements.
2. The second step had the aim of estimating error. The usual approach consists in the computation of the difference between the measured load and the applied one. In the case of multiple loads acting simultaneously, the norm of the difference between the two load vectors is estimated. A drawback of this approach consists in the error being estimated in absolute terms. Its application would be impossible in the present case, as the vector of loads consists in both forces and moments, which have different measuring units and cannot be summed up together. Therefore, the error was computed as in Eq. 2.86, where the square root of the sum of the squares of relative errors is computed.

$$\begin{aligned}
 \text{Error} = & \left[\beta_1 \cdot \left(\frac{F_{x_m} - F_{x_a}}{F_{x_a}} \right)^2 + \beta_2 \cdot \left(\frac{F_{y_m} - F_{y_a}}{F_{y_a}} \right)^2 \right. \\
 & + \beta_3 \cdot \left(\frac{F_{z_m} - F_{z_a}}{F_{z_a}} \right)^2 + \beta_4 \cdot \left(\frac{M_{x_m} - M_{x_a}}{M_{x_a}} \right)^2 \\
 & \left. + \beta_5 \cdot \left(\frac{M_{y_m} - M_{y_a}}{M_{y_a}} \right)^2 + \beta_6 \cdot \left(\frac{M_{z_m} - M_{z_a}}{M_{z_a}} \right)^2 \right]^{\frac{1}{2}} \quad (2.86)
 \end{aligned}$$

Writing the error in this form was also the opportunity to incorporate additional weight coefficients, β_1, \dots, β_6 , which take the importance of the measured component into account. For instance, they can all be kept at a unitarian value, except β_1 , to account for the minor importance of a very accurate estimation of F_x .

The numerical procedure for the determination of the most suitable value of the *damping coefficient* can be summarized as follows.

- Just one of the sample strain readings was initially considered. According to the described methodology to simulate experimental uncertainty, 500 random readings were determined. The value of α was initially assumed as unitarian and the model in Eqs. 2.84 and 2.85 was followed for the computation of the vector of loads, \mathbf{f} , corresponding to the first of the 500 random input readings.
- Afterwards, the error was computed by the application of Eq. 2.86, where the subscript m stands for *measured* and the subscript a stands for *applied*. The same procedure was followed for the remaining 499 random input signals, thus computing other error terms.
- An averaged value of error was finally determined for $\alpha = 1$ and recorded. Then the value of α was halved and the whole procedure was followed again for the 500 random signals, thus computing a further value of the averaged error for $\alpha = 0.5$. Then the value was halved again the and the entire procedure was followed until it assumed a value around 10^{-3} .

The orders of magnitude seemed quite reasonable to define the interval where the most suitable value of α was likely to be.

- The average error versus the number of the performed iteration $It.$ is plotted in Fig. 2.55. The simple relationship between the iteration $It.$ and the actual value of α is shown in Eq. 2.87.

$$\alpha_{It.} = \left(\frac{1}{2}\right)^{It.-1} \tag{2.87}$$

- It can be observed that the curve exhibits a minimum for $It. = 7$, corresponding to $\alpha = 0.016$. It suggests that the performed procedure is really effective at reducing the error induced by the experimental uncertainty, thus improving device robustness.
- A possible drawback lies in the determined value depending on the sample reading used to generate the 500 random input signals. For this reason the whole procedure was repeated also for other sample readings. The results were all consistent and proved that the most suitable value for α is always close to 0.016.

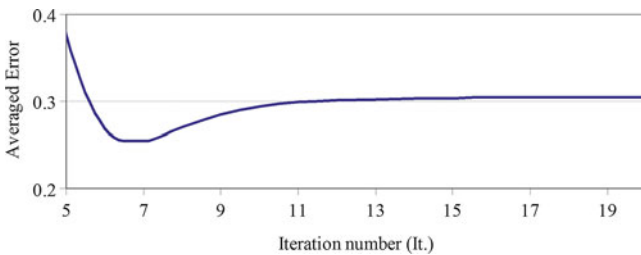


Fig. 2.55 Average error as a function of the iteration number ($It.$)

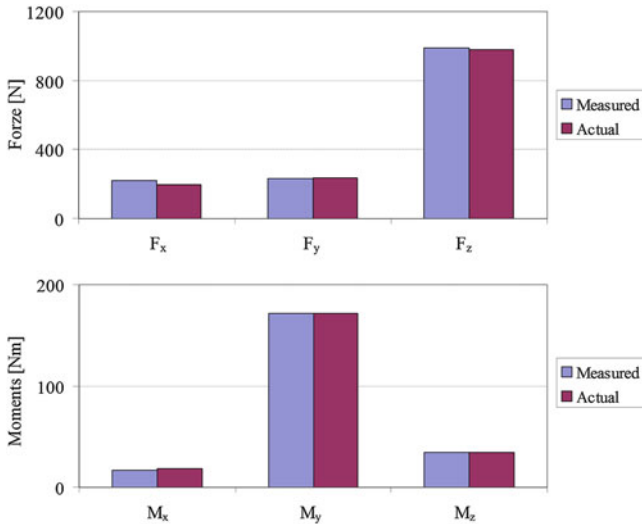


Fig. 2.56 Comparison between actual and measured force and moments after regularization

The final issue concerns the impact of the two strategies that have been tackled and described here, to account for the offset terms and to achieve a regularization of the ill-problem. The histograms in Fig. 2.56 compare the estimated loads to the current applied loads in case of strain readings affected by a 5% error. These results, to be compared with those of Fig. 2.52, denote a very good agreement with a maximum acceptable discrepancy around 11% for F_x .

References

1. Tokyo sokki kenkyujo strain gages (2010). https://www.tml.jp/e/download/catalog/STRAIN_GAUGES.pdf
2. Zhonhhang electronic measuring instruments (2010). <http://www.zemiceurope.com/media/Straingagescatalogue.pdf>
3. (2014) <http://www.ni.com/white-paper/3642/en/>
4. (2014) <http://www.ni.com/white-paper/4176/en/>
5. Strain gauges: absolute precision from HBM (HBM strain gauge catalogue) (2014). Available on-line at <http://www.hbm.com/fileadmin/mediapool/hbmdoc/technical/s1265.pdf>
6. Vishay(micromesur.) strain gages (2014). <http://vishaypg.com/micro-measurements/stress-analysis-strain-gages/>
7. Ajovalasit A (2008) Analisi sperimentale delle tensioni con gli estensimetri elettrici a resistenza, vol 1. Aracne Roma, Roma
8. Bray A, Vicentini V (1975) Meccanica sperimentale: misure ed analisi delle sollecitazioni. Levrotto & Bella, Turin
9. Dally JW, Riley WF (2005) Experimental stress analysis, 4th edn. McGraw-Hill, NY
10. Hoffmann K (1989) An introduction to measurements using strain gauges. Hottinger Baldwin Messtechnik GmbH, Darmstadt

11. Kobayashi AS (1993) Handbook on experimental mechanics. SEM
12. Sharpe WN Jr, Sharpe WN (2008) Springer handbook of experimental solid mechanics. SEM, Springer
13. Kyowa strain gages (2010). http://www.kyowa-ei.com/eng/product/category/strain_gages/kfg/index.html
14. (2014) <http://www.tutelman.com/aboutus/index.html>
15. Figliola R, Beasley D (2011) Theory and design for mechanical measurements, 5th edn. Wiley, Hoboken
16. Machin KE, Hassan YE (1976) Transverse-sensitivity errors in strain-gage measurement. *Exp Mech* 16(1):38–40
17. Vishay (2011) Technical note TN-509, Errors due to transverse sensitivity in strain gages. <http://www.vishaypg.com/docs/11059/tn509tn5.pdf>
18. How is temperature affecting your strain measurement accuracy? (2011). <http://www.ni.com/white-paper/3432/en/>
19. Doyle JF (2004) Modern Experimental stress analysis. Completing the solution of partially specified problems, vol 1, Wiley. ISBN 0-470-86156-8
20. Vishay (2010) Technical note TN-612, Three-wire quarter bridge circuit. <http://www.vishaypg.com/docs/11092/tt612.pdf>
21. Olmi G (In press) Failure of the chassis of roller skates for agonistic figure skating. *Case Stud Eng Fail Anal*. doi:10.1016/j.csefa.2014.08.002
22. De Cesare G, Gavesi M, Palma F, Riccò B (2003) A novel a-Si:H mechanical stress sensor. *Thin Solid Films* 427:191–195
23. Pramanik C, Islam T, Saha H (2006) Temperature compensation of piezoresistive micro-machined porous silicon pressure sensor by ANN. *Microelectron Reliab* 46:343–351
24. Olmi G (2009) Investigation on the influence of temperature variation on the response of miniaturized piezoresistive sensors. *Strain* 45(1):63–76
25. Ballo F, Gobbi M, Mastinu G, Previati G (2014) Advances in force and moments measurements by an innovative six-axis load cell. *Exp Mech* 54(4):571–592
26. Boyd T, Hull ML, Wootten D (1996) An improved accuracy six-load component pedal dynamometer for cycling. *J Biomech* 29:1105–1110
27. Chao LP, Chen KT (1997) Shape optimal design and force sensitivity evaluation of six-axis force sensors. *Sens Actuators A: Phys* 63:105–112
28. Chao LP, Yin CY (1999) The six-component force sensor for measuring the loading of the feet in locomotion. *Mater Des* 20:237–244
29. Freddi A, Olmi G, Croccolo D (2002) Ein neuer Aufnehmer zur direkten Messung der von einem Skifahrer auf die Skier übertragenen Kräfte und Momente. *Messtechnische Briefe (MTB)* 1(2002):3–10
30. Gobbi M, Previati G, Guarneri P, Mastinu G (2011) A new six-axis load cell. Part II: error analysis, construction and experimental assessment of performances. *Exp Mech* 51:389–399
31. Gorinevsky D, Formalsky A, Schneider A (1998) Force control of robotics systems. CRC Press, New York
32. Kim GS, Kang DI, Ree SH (1999) Design and fabrication of a six-component force/moment sensor. *Sens Actuators A: Phys* 77:209–220
33. Mastinu G, Gobbi M, Previati G (2011) A new six-axis load cell. Part I: design. *Exp Mech* 51:373–388
34. Olmi G (2012) A novel method for strain controlled tests. *Exp Mech* 52(4):379–393
35. Olmi G (2014) An experimental investigation on a crack propagating from a geartrain housing in an asphalt milling machine. *Eng Fail Anal* 38:38–48
36. Olmi G, Freddi A, Croccolo D (2008) In-field measurement of forces and deformations at the rear end of a motorcycle and structural optimisation: experimental-numerical approach aimed at structural optimization. *Strain* 44(6):453–461
37. Pelliccioli G, Petrone N (2010) Experimental strain analysis of the mast of a 420 sailboat during sailing. *Strain* 46:482–492

38. Petrone N (2011) Acquisition and analysis of ground reaction forces and foot orientation of in-line skates during track speed skating. *Proc XXI Congr, Int Soc Biomech, J Biomech*, 398
39. Petrone N (2011) Field acquisition of ski boots flexural behaviour for the correct definition of standard laboratory tests. In: *Proceedings 19th international conference on Ski Trauma and Safety*, Keystone, Colorado, United States
40. Scott NW, Yoneyama T, Kagawa H, Osada K (2007) Measurement of ski snow-pressure profiles. *Sports Eng* 10:145–156
41. Blaysat B, Florentin E, Lubineau G, Moussawi A (2012) A dissipation gap method for full-field measurement-based identification of elasto-plastic material parameters. *Int J Numer Methods Eng* 91:685–704
42. Florentin E, Lubineau G (2010) Identification of the parameters of an elastic material model using the constitutive equation gap method. *Comput Mech* 46:521–531
43. Lubineau G (2009) A goal-oriented field measurement filtering technique for the identification of material model parameters. *Comput Mech* 44:591–603
44. Lubineau G, Florentin E, Moussawi A, Blaysat B (2012) Une approche en dissipation pour l'identification de propriétés matériaux hétérogènes à partir de mesures de champs. *Matériaux & Techniques* 100:665–670
45. Moussawi A, Lubineau G, Florentin E, Blaysat B (2013) The constitutive compatibility method for identification of material parameters based on full-field measurements. *Comput Methods Appl Mech Eng* 265:1–14
46. Pierson J (2014) Load/force cells: how they work. Available on-line at <http://www.sensorland.com/HowPage005.html#anchor619131>
47. Rumul testronic (2014). http://www.rumul.ch/pdf/testronic_e.pdf
48. Jung P (1975) Zur messtechnik an skiern. *Messtechnische Briefe* 11(2):34–37
49. Jung P (1975) Zur messtechnik an skiern. *Messtechnische Briefe* 11(3):59–62
50. Gradassi AR (2001) Progetto e costruzione di una cella di carico per la misurazione di forze in attrezzature sportive (scarpone da sci). MS thesis, University of Bologna
51. Olmi G (2001) Progetto e costruzione di una cella di carico di seconda generazione per la misurazione di forze in attrezzature sportive. MS thesis, University of Bologna
52. Laermann KH (2008) *Inverse problems in experimental structural analysis*. Shaker Verlag, Aachen
53. Liu GR, Han X (2003) *Computational inverse techniques in nondestructive evaluation*. CRC Press LLC, Boca Raton
54. Tikhonov A, Goncharksky A, Stepanov V, Yagola A (1995) *Numerical methods for the solution of ill-posed problems*. Kluwer, Dordrecht
55. Akbari S, Taheri-Behrooz F, Shokrieh M (2013) Slitting measurement of residual hoop stresses through the wall-thickness of a filament wound composite ring. *Exp Mech* 53(9):1509–1518
56. Jones IA, Wang P, Becker AA (2011) Constrained least-squares solution and regularization in inverse boundary element analysis of photoelastic models involving contact. *J Strain Anal Eng Des* 46(8):709–726
57. Kretschmar J, Stockmann M, Shutov A (2011) Determination of pressure distributions using a gradient based optimization method and application at forming tools for high gears. In: *Proceedings 10th youth symposium on experimental solid mechanics*. Chemnitz University of Technology, Institute of Mechanics and Thermodynamics, Department of Solid Mechanics, Chemnitz, pp 65–66
58. Golub GH, Heath M, Wahba G (1979) Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* 21(2):215–223
59. Hansen PC (1993) Analysis of discrete ill-posed problems by means of the l-curve. *SIAM Rev* 34:561–580
60. Hansen PC (1994) Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems. *Numer Algorithms* 6:1–35
61. Hansen PC, O'Leary DP (1993) The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM J Sci Comput* 14(6):1487–1503

62. Lawson CL, Hanson RJ (1974) Solving least squares problems. Prentice-Hall, Englewood Cliffs
63. Olmi G (In press) An efficient approach to ill-posed problem regularization applied to an over-determined six-degree of freedom load cell. Exp Mech. doi:[10.1007/s11340-015-9986-3](https://doi.org/10.1007/s11340-015-9986-3)

Chapter 3

Introduction to Photoelasticity

Abstract For nearly a century the photoelastic method has been a reference to verify special applications of the theory of elasticity and to provide a solution to cases of loaded structures without theoretical models. Today photoelasticity has been largely replaced by numerical methods, such as finite elements, which provide solutions for any structural problem with great detail and accuracy, and is therefore confined to the role of an illustration tool. This technique is based on an analogy of the optical behavior of transparent amorphous bodies; this fact limits the applications but its capability of analyzing the general field and, at the same time, clarifying details, is still very useful, and applications of photoelastic techniques continue to be presented in the most qualified international venues. It is therefore still logical to present an overview of the theory and to discuss some applications. The classical approach remains the best way of understanding the inherent advantages and drawbacks and, thanks to the inverse approach, continues to be a useful investigation method.

3.1 Premise

Recently, photoelasticity has achieved significant improvements for a digital extension (digital photoelasticity), developed and applied by several researchers [1–5].¹

3.2 Nature of Light

According to the electromagnetic theory due to Maxwell, light is a wave phenomenon related to the harmonic oscillation of an electric field (which can be represented by one or more vectors normal to the direction of light propagation) and to harmonic oscillation of a magnetic field perpendicular to the first [6].

¹ The Society of Experimental Mechanics SEM in the United States included about 80 papers in the annual 2009 conference and comparable numbers in previous years. At the ICEM (International Conference of Experimental Mechanics in 2014) two sessions were dedicated to this method.

Let us first suppose that we have a point source of monochromatic light and consider only the electric field. Since this field oscillates harmonically, its value at a given instant of time t and at a point at a certain distance r from the source S can be expressed as (Figs. 3.1 and 3.2):

$$a(r, t) = \frac{K}{r} \cos \left[\frac{2\pi}{\lambda} (r - ct) + \phi \right] \tag{3.1}$$

where:

- K/r = amplitude or attenuation coefficient
- λ = wavelength
- r = distance of the point from the source
- t = time variable
- c = light propagation velocity in the medium
- $\left[\frac{2\pi}{\lambda} (r - ct) + \phi \right]$ = phase angle (or phase)
- ϕ = initial phase angle (or initial phase) for $t = 0$ and $r = 0$.

If $t = \text{const.}$, $r = \text{const.}$, $\phi = \text{const.}$, $a(r, t)$ is constant on a spherical surface with center S . At a distance r from the source, the electric field has the same value as the phase angle in any direction.

The propagation front is spherical, i.e. the oscillatory phenomenon is in phase at all the points on a spherical surface normal to the propagation direction. There are other forms of wave fronts: a plane wavefront can be obtained placing a point source in the focus of a plane-convex lens.

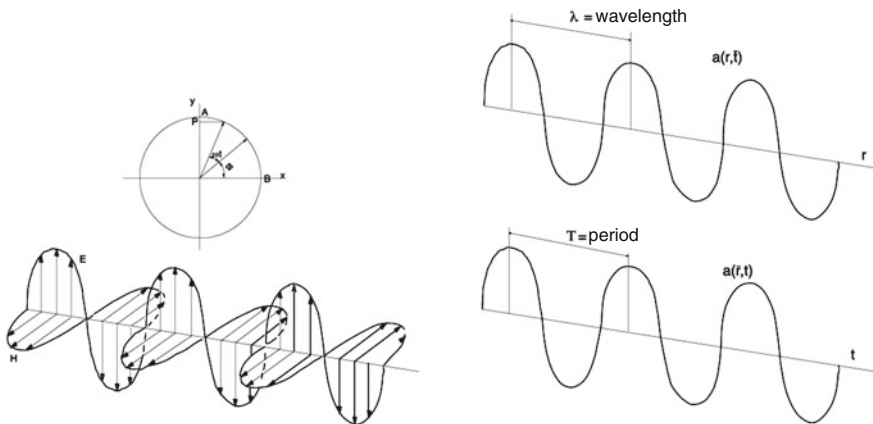
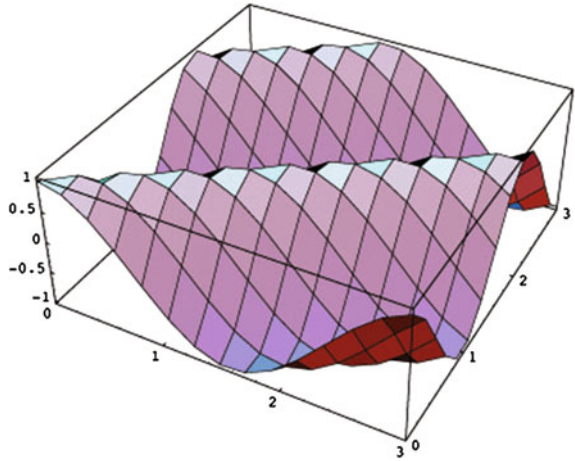


Fig. 3.1 Oscillation of electric field with associated magnetic field on *left side* and representation of the function $a(r, t)$ on the *right side*

Fig. 3.2 Three-dimensional representation of $a(r, t)$ in coordinates r and t



The instantaneous value of $a(r, t)$ can also be expressed as a real part of a complex vector, indicated by the capital boldface letter \mathbf{A} :

$$a(r, t) = \text{Re} [\mathbf{A}]$$

where:

$$\underline{\mathbf{A}} = \frac{K}{r} e^{i \left[\frac{2\pi}{\lambda} (r-ct) + \phi \right]} = \frac{K}{r} e^{i\phi} \cdot e^{i \left(\frac{2\pi}{\lambda} r - \omega t \right)} \quad (3.2)$$

If $f = c/\lambda$ is the oscillation frequency of the electric field and $\omega = 2\pi f$ the corresponding pulsation, Eq. 3.1 can be interpreted as the combination of an harmonic vibration and a uniform translation.

If t is treated as a constant, Eq. 3.1 gives the wave profile along r at this instant; vice-versa Eq. 3.1 for a constant value $r = \bar{r}$ gives the oscillation at distance r from the origin.

The wavelength λ is defined for variable r and period $T = \frac{\lambda}{c}$ or frequency $f = \frac{1}{T}$ are defined for variable t .²

Denoting by $|A|$ the amplitude K/r and $\mathbf{A} = \frac{K}{r} e^{i\phi} = |A| e^{i\phi}$ as complex amplitude, Eq. 3.2 can be written as (Fig. 3.2):

$$a(r, t) = \text{Re} \left[\mathbf{A} e^{i \left(\frac{2\pi}{\lambda} r - \omega t \right)} \right] \quad (3.3)$$

² The light color is determined by the wavelength; the values of the visible spectrum range from dark red ($f = 390 \times 10^{12}$ Hz or $\lambda = 770$ nm) to violet ($f = 770 \times 10^{12}$ Hz, or $\lambda = 390$ nm) ($1 \text{ nm} = 10^{-9} \text{ m}$), passing through orange, yellow, green, blue, indigo-violet. If the light vector is composed of vibrations a_1, a_2, \dots, a_n of the same frequency, light is monochromatic; if the components have different frequencies, the eye perceives them all together as white light.

3.2.1 Polarized Light

A wave of monochromatic light is said to be *polarized* when it is represented by vectors with non-random orientations.

The electric field, i.e. the *Light vector* in a point of the propagation path, is completely determined by a complex amplitude that can be represented by its projections on the coordinate axes. i.e. by means of two plane polarized waves on coordinate planes xr , yr with complex amplitudes:

$$|A_x|e^{i\phi_x}$$

$$|A_y|e^{i\phi_y}$$

and the light vector can be written as:

$$\mathbf{A} = \begin{pmatrix} |A_x|e^{i\phi_x} \\ |A_y|e^{i\phi_y} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_x \\ \mathbf{A}_y \end{pmatrix} \quad (3.4)$$

If ϕ_x and ϕ_y are expressed as functions of their sum $\beta = \phi_x + \phi_y$ and their difference $\Delta = \phi_y - \phi_x$, the following relationships hold:

$$\begin{cases} \phi_x = \frac{\beta - \Delta}{2} \\ \phi_y = \frac{\beta + \Delta}{2} \end{cases}$$

and Eq. 3.4 can be written as:

$$\begin{pmatrix} |A_x|e^{i\phi_x} \\ |A_y|e^{i\phi_y} \end{pmatrix} = \begin{pmatrix} |A_x| \\ |A_y|e^{i\Delta} \end{pmatrix} e^{i\beta/2} e^{-i\Delta/2}$$

Terms such as $e^{i\beta/2} e^{-i\Delta/2}$ are common phases to both vectors and can be omitted in the following discussion of photoelasticity, sensitive only to phase differences between the two vectors.

3.2.2 Plane Polarized Light

If this is the phase relationship between the normal coordinates components:

$$\Delta = \begin{cases} 0 \\ \pi \end{cases} \quad (3.5)$$

the resultant electric field is plane polarized light (Fig. 3.3). For $\Delta = 0$ and $\Delta = \pi$ is:

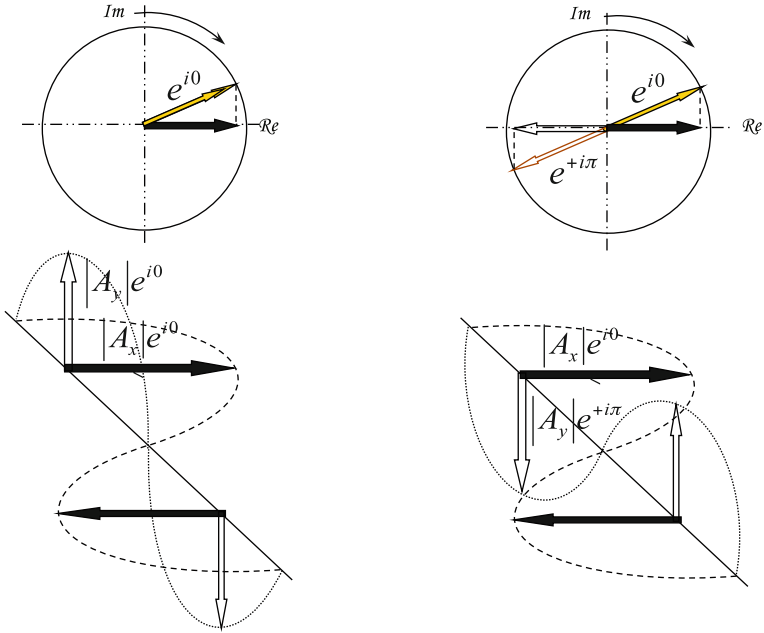


Fig. 3.3 Plane polarized light

$$\mathbf{A} \equiv \begin{pmatrix} |A_x| \\ |A_y|e^{i0} \end{pmatrix} = \begin{pmatrix} |A_x| \\ +|A_y| \end{pmatrix} \tag{3.6}$$

$$\mathbf{A} \equiv \begin{pmatrix} |A_x| \\ |A_y|e^{i\pi} \end{pmatrix} = \begin{pmatrix} |A_x| \\ -|A_y| \end{pmatrix}$$

In general, for $|A_x| \neq |A_y|$ the vibration plane is inclined with respect to the x axis of an angle α , such that:

$$\tan \alpha = \frac{a_y}{a_x} = \frac{|A_y|}{|A_x|} = \text{const.}$$

If also $|A_x| = |A_y| = |A|$ the light vector becomes:

$$A = |A| \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}$$

If $\tan \alpha = \pm 1$, $\alpha = \pm \frac{\pi}{4}$ and light³ is plane polarized at 45° .

Polarized light in planes xr and yr is represented by vectors:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} |A_x|$$

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} |A_y|$$

3.2.3 Circularly Polarized Light

A wave of monochromatic light is said to be circularly polarized when the electric field vector is such that:

$$|A_x| = |A_y| = |A| \quad \text{and} \quad \Delta = \pm\pi/2$$

The vector of the resulting electric field can be expressed in the following way, except for a common phase which does not influence the light properties.

$$\mathbf{A} \equiv |A| \begin{pmatrix} 1 \\ e^{\pm i\pi/2} \end{pmatrix} \equiv |A| \begin{pmatrix} 1 \\ \pm i \end{pmatrix} \quad (3.7)$$

Instant field values can be deduced from the instant components values along x and y . For Eq. (3.1), taking Eq. 3.7 into account, they are:

$$\begin{cases} a_x = |A| \cos \left[\frac{2\pi}{\lambda}(r - ct) + \phi_x \right] \\ a_y = \mp |A| \sin \left[\frac{2\pi}{\lambda}(r - ct) + \phi_x \right] \end{cases} \quad (3.8)$$

Squaring and summing the instantaneous values of both components:

³ The instantaneous value of the field in planes xr and yr is expressed, for Eq. 3.1, by:

$$a_x = |A_x| \cos \left[\frac{2\pi}{\lambda}(r - ct) + \phi_x \right]$$

$$a_y = \pm |A_y| \cos \left[\frac{2\pi}{\lambda}(r - ct) + \phi_x \right]$$

and instantaneous value of the resulting vector is:

$$a_r = \sqrt{a_x^2 + a_y^2} = \sqrt{2}|A| \cos \left[\frac{2\pi}{\lambda}(r - ct) + \phi_x \right].$$

$$a_r = \sqrt{a_x^2 + a_y^2} = |A| \quad (3.9)$$

an constant instant value equal to $|A|$ is obtained. The angle the light vector forms with x axis is given by:

$$\tan \alpha = \frac{a_y}{a_x} = \mp \tan \left[\frac{2\pi}{\lambda}(r - ct) + \phi_x \right] \quad (3.10)$$

i.e.:

$$\alpha = \mp \frac{2\pi}{\lambda}(r - ct) + \phi_x$$

The angle α varies linearly as a function of time t and r , while the vector amplitude is constant. The vector describes a circle with the variable t with an angular velocity $s\omega = \frac{2\pi}{\lambda} c$. Since it varies in position along the propagation direction, the composed movement is an elliptical helix.

The sign distinguishes the direction of rotation of the vector. For an observer looking at the source, the sign is assumed positive when the rotation is clockwise and vice-versa for counterclockwise. With this convention, when t increases, angle α decreases.

3.2.4 Elliptically Polarized Light

Light is elliptically polarized with the ellipse diameters along x and y axes, when $A_x \neq A_y$ and Δ is equal to $\pm\pi/2$.

The electric field vector can be expressed as:

$$\begin{pmatrix} |A_x| \\ \pm i |A_y| \end{pmatrix} \quad (3.11)$$

and along x and y :

$$\begin{cases} a_x = |A_x| \cos \left[\frac{2\pi}{\lambda}(r - ct) + \phi_x \right] \\ a_y = \mp |A_y| \sin \left[\frac{2\pi}{\lambda}(r - ct) + \phi_x \right] \end{cases} \quad (3.12)$$

Squaring and summing:

$$\frac{a_x^2}{|A_x|^2} + \frac{a_y^2}{|A_y|^2} = 1 \quad (3.13)$$

If $|A_x| \neq |A_y|$ and $\phi_y - \phi_x \neq \pi/2$ light is still elliptically polarized even if with axes not coinciding with the coordinate axes. See a summary of all these cases in Fig. 3.4.

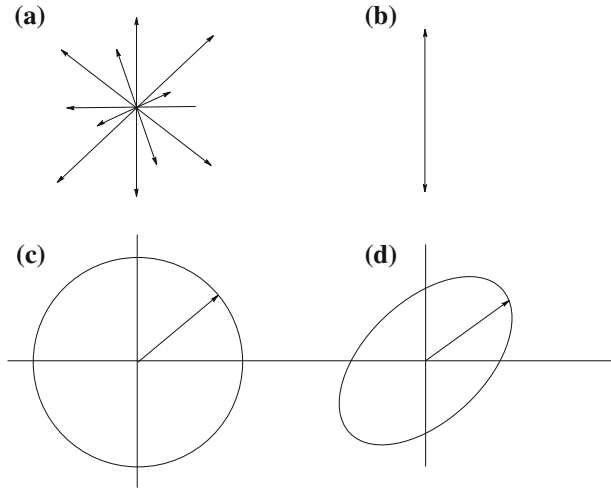


Fig. 3.4 Polarized light. **a** Natural light, **b** plane polarized, **c** circularly polarized, **d** elliptically polarized)

3.2.5 Plane Polarizer

In order to alter the state of light polarization, different types of polarizing filters can be set along light beams.

Plane polarizers are optical filters that allow only vectors components vibrating in a defined direction, called *polarizer axis* to pass through. Components along its normal direction are totally absorbed.

Identifying light vectors with the corresponding complex amplitudes, and considering only a wavelength, the incident light is given by, Fig. 3.5:

$$\begin{pmatrix} \mathbf{A}_x \\ \mathbf{A}_y \end{pmatrix}$$

The angle formed by *P* axis with *x* axis is ϑ (considered positive in counterclockwise direction, looking at the light source).

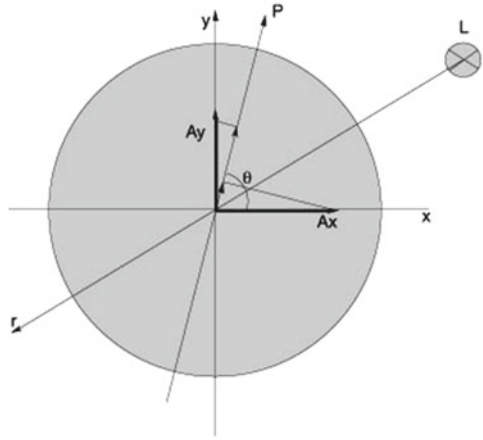
The emerging light vector, referring to the *P* axis and to its perpendicular direction, is:

$$\mathbf{A}'_p = \mathbf{A}_x \cos \vartheta + \mathbf{A}_y \sin \vartheta \tag{3.14}$$

If the emerging beam refers to cartesian axes, it becomes:

$$\begin{pmatrix} \mathbf{A}'_x \\ \mathbf{A}'_y \end{pmatrix} = \begin{pmatrix} \cos^2 \vartheta & \sin \vartheta \cos \vartheta \\ \sin \vartheta \cos \vartheta & \sin^2 \vartheta \end{pmatrix} \cdot \begin{pmatrix} \mathbf{A}_x \\ \mathbf{A}_y \end{pmatrix} \tag{3.15}$$

Fig. 3.5 L light source, P polarizer axis, x and y Cartesian coordinates



Matrix \mathbf{P}_ϑ which transforms the incident light vector into an emerging light vector is called a *polarizer matrix*.⁴

If the polarizer axis is respectively along the x axis ($\vartheta = 0$), y axis ($\vartheta = \pi/2$) or at 45° ($\vartheta = \pi/4$), \mathbf{P} assumes respectively the following forms:

$$\mathbf{P}_x = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \mathbf{P}_y = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \mathbf{P}_{\pi/4} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad (3.16)$$

3.2.6 Double Refraction or Birefringence

Some transparent materials have the property of splitting the light that passes through them in two polarized components, lying on two planes perpendicular to each other; these components are transmitted with different velocities. The phenomenon is called *double refraction or birefringence*.

Referring to Fig. 3.6, let be x, y be the coordinate axes and 1, 2 the polarization axis of the plate.

⁴ Currently the employed filters are *Polaroid®* (Polaroid Co., Cambridge, Mass, USA), supplied in sheets of various size; “this material, known as J-sheet, was later replaced by the improved H-sheet. H-sheet is a polyvinyl alcohol (PVA) polymer impregnated with iodine. During manufacture, the PVA polymer chains are stretched such that they form an array of aligned, linear molecules in the material. The iodine dopant attaches to the PVA molecules and makes them conducting along the length of the chains. Light polarized parallel to the chains is absorbed, and light polarized perpendicular to the chains is transmitted” [7].

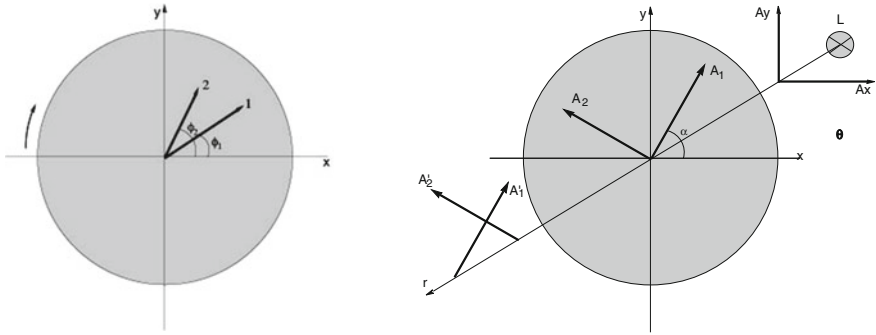


Fig. 3.6 Phase shift introduced by a birefringent filter in complex representation on the *left* and in physical space on the *right* ($\Delta = \Phi_2 - \Phi_1$)

Axis 1 forms an angle α respect to axis x (with the same sign convention). The incident light vector

$$\begin{pmatrix} \mathbf{A}_x \\ \mathbf{A}_y \end{pmatrix}$$

referring to the polarizing axes 1 and 2 becomes:

$$\begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix} = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} \cdot \begin{pmatrix} \mathbf{A}_x \\ \mathbf{A}_y \end{pmatrix} \tag{3.17}$$

While components \mathbf{A}_1 and \mathbf{A}_2 are at the filter entrance, the following ones are emerging:

$$\begin{pmatrix} \mathbf{A}'_1 \\ \mathbf{A}'_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\Delta} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix} \tag{3.18}$$

Referring them to x and y is:

$$\begin{pmatrix} \mathbf{A}'_x \\ \mathbf{A}'_y \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \cdot \begin{pmatrix} \mathbf{A}'_1 \\ \mathbf{A}'_2 \end{pmatrix} \tag{3.19}$$

From Eq. 3.20 the following is derived:

$$\begin{pmatrix} \mathbf{A}'_x \\ \mathbf{A}'_y \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & e^{i\Delta} \end{pmatrix} \cdot \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} \cdot \begin{pmatrix} \mathbf{A}_x \\ \mathbf{A}_y \end{pmatrix} \tag{3.20}$$

From the matrix product:

$$\mathbf{M} = \begin{pmatrix} \cos^2 \alpha + e^{i\Delta} \sin^2 \alpha & (1 - e^{i\Delta}) \sin \alpha \cos \alpha \\ (1 - e^{i\Delta}) \sin \alpha \cos \alpha & \sin^2 \alpha + e^{i\Delta} \cos^2 \alpha \end{pmatrix} \quad (3.21)$$

Synthetically:

$$\mathbf{A}' = \mathbf{M} \cdot \mathbf{A} \quad (3.22)$$

If:

$$\begin{cases} m_{11} = \cos^2 \alpha + e^{i\Delta} \sin^2 \alpha \\ m_{12} = m_{21} = (1 - e^{i\Delta}) \sin \alpha \cos \alpha \\ m_{22} = \sin^2 \alpha + e^{i\Delta} \cos^2 \alpha \end{cases} \quad (3.23)$$

then:

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \quad (3.24)$$

When phase difference is equal to $\pi/2$ the birefringent filter is called a *quarter-wave plate*, because the phase difference between the two polarized rays at the filter exit is equal to one-quarter of the wave length for the considered monochromatic light.

If the phase shift is π and 2π , the filter is called respectively *half-wave* and *full-wave* plate.

If $\Delta = \frac{\pi}{2}$ and $\alpha = \pm \frac{\pi}{4}$ the matrix of the birefringent filter becomes:

$$\mathbf{Q} = \frac{1}{2} \begin{pmatrix} 1+i & \pm(1-i) \\ \pm(1-i) & 1+i \end{pmatrix} e^{i\phi_1} \quad (3.25)$$

Collecting the common term $(1+i)$ (equal to $2e^{i\pi/4}$) that together with vector $e^{i\phi_1}$ do not enter, as it will be seen, in the interference phenomenon, the matrix \mathbf{Q} can be expressed, except for these two vectors, as:

$$\begin{pmatrix} 1 & \mp i \\ \mp i & 1 \end{pmatrix} \quad (3.26)$$

3.3 Light Treatment in an Optical System

Let now examine the behavior of light passing through an optical system composed of elements designed to alter polarization and the phase of the light. For the linearity property of transformations that the light undergoes, linear matrix algebra can be introduced [8] that makes it possible to express the emerging field as a function of the field at entry. Each element of the optical system is represented by a matrix. For the property of matrix algebra, the field emerging from the optical system is obtained

by pre-multiplying the incident field by the matrices of the optical elements taken in reverse order:

$$\mathbf{A}' = \mathbf{M}_n, \dots, \mathbf{M}_i, \dots, \mathbf{M}_2, \mathbf{M}_1 \cdot \mathbf{A} \quad (3.27)$$

All the matrices \mathbf{M}_i are square, of the second order and symmetric.

3.3.1 Optical System with a Polarizer and a Birefringent Filter

Let us consider the generation of different kinds of polarized light, setting up an optical system constituted by a polarizer filter followed by a birefringent one.

Plane, circular or elliptical polarized lights can be generated depending on phase shift Δ and on angle α .

3.3.1.1 Circular Polarized Light

The simplest case is given by the following conditions:

$$\begin{cases} \vartheta = 0 & \text{and } \alpha = 0 \text{ or } \alpha = \pi/2 \\ \vartheta = +\frac{\pi}{2} & \text{and } \alpha = 0 \text{ or } \alpha = \pi/2 \end{cases}$$

For any value of Δ , the emerging light remains plane polarized in the horizontal or in the vertical plane.

Setting up a polarizer with polarizing axis along y followed by a Quarter wave plate, Fig. 3.7, respectively for $\alpha = \pi/4$ and for $\alpha = -\pi/4$, the emerging vector becomes:

$$\mathbf{A} = \begin{pmatrix} 1 & \mp i \\ \mp i & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{A}_y = \begin{pmatrix} \mp i \\ 1 \end{pmatrix} \mathbf{A}_y \equiv \begin{pmatrix} \mp 1 \\ i \end{pmatrix} \mathbf{A}_y \quad (3.28)$$

having collected and neglected $i = e^{i\pi/2}$, the absolute value of a phase shift that is also irrelevant in the interference phenomenon.

The emerging light is circularly polarized, Eq. 3.7 and Fig. 3.8.

3.3.1.2 Elliptic Polarized Light

If the filter is a Quarter-wave plate, for any value of α , but $0, \pi/2, \pi/4$ (or even multiples of them), the transmitted light is elliptically polarized.

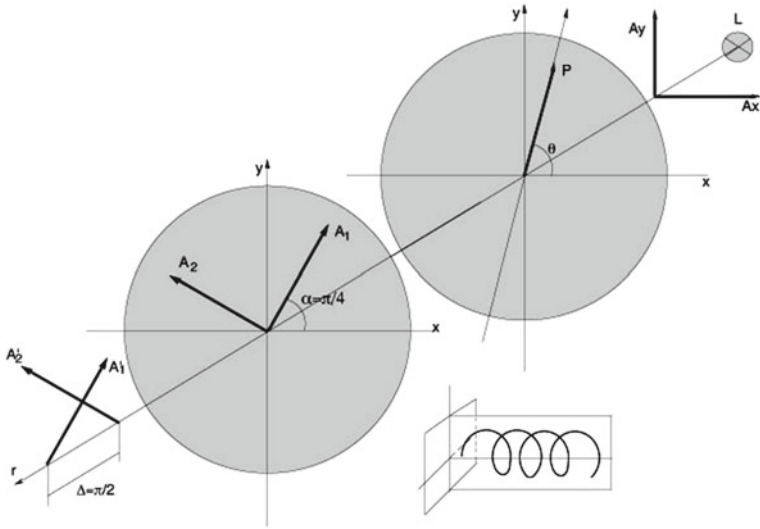


Fig. 3.7 Light treatment with a polarizer and a quarter-wave plate

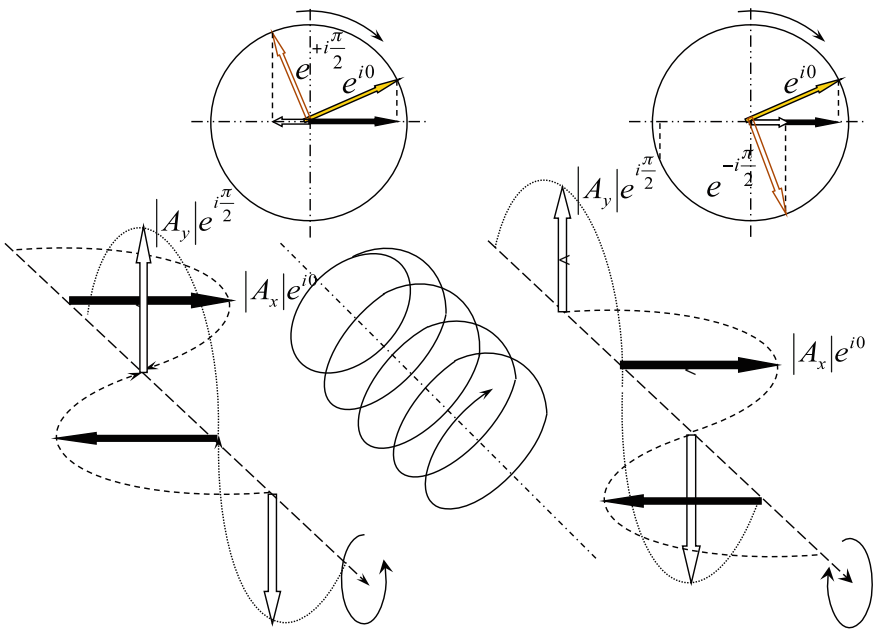


Fig. 3.8 Generation of circular polarized light

3.4 Light Intensity

The intensity of the light radiation, which is responsible for common lighting effects (photochemical reactions, stimulation of the retina, etc.) is proportional to the square of the amplitude modulus.

$$I \propto |A^2| \quad (3.29)$$

Introducing the complex notation Eq. 3.29, it is possible to write also:

$$I \propto |A|e^{-i\phi}|A|e^{i\phi} = A^*A \propto |A|^2 \quad (3.30)$$

that is the product of amplitude by its conjugate (A^* being the conjugate of A).

More generally, if an electric field is expressed by:

$$\begin{pmatrix} |A_x|e^{i\phi_x} \\ |A_y|e^{i\phi_y} \end{pmatrix}$$

light intensity is represented by the scalar:

$$I \propto (|A_x|e^{-i\phi_x} \ |A_y|e^{-i\phi_y}) \cdot \begin{pmatrix} |A_x|e^{i\phi_x} \\ |A_y|e^{i\phi_y} \end{pmatrix} = |A_x|^2 + |A_y|^2 \quad (3.31)$$

i.e.

$$I \propto \mathbf{A}^* \cdot \mathbf{A} \quad (3.32)$$

where \mathbf{A}^* is the transpose vector of the conjugate of \mathbf{A} .

An absolute phase does not contribute to light intensity, i.e. to energy transport.

3.5 Optics of Photoelasticity

Many non-crystalline transparent materials [6, 9–15], ordinarily optically isotropic, if loaded, become anisotropic and show a behavior similar to crystals. This effect, that vanishes when loads are removed, is called *temporary or artificial birefringence*.

When this happens, the transparent material has *photoelastic* behavior and is called photoelastic material.

Let us consider a plane model of photoelastic material. Initially the model is unloaded and has a refractive index equal to n_0 at every points and in any direction. When the model is loaded, in any point it becomes birefringent, like the optical filter previously seen.

The principal optical directions with maximum and minimum refractive indices coincide with the principal directions 1 and 2 of the stress state in that point; phases of the two vectors along 1 and 2 axes also vary along the model thickness so that

both components emerge at different times. If n_1 and n_2 are refractive indices at the point along axes 1 and 2 and d is the thickness, the delay of A'_2 on A'_1 at the exit is:

$$t_2 - t_1 = \frac{d}{v_2} - \frac{d}{v_1} = \frac{d}{c} \left(\frac{c}{v_2} - \frac{c}{v_1} \right) = \frac{d}{c} (n_2 - n_1) \quad (3.33)$$

where c is the speed of light in vacuum.

The axis of polarization 1 with a smaller refractive index n_1 , thus with greater propagation speed, is called a *fast axis*, while the other is called a *slow axis*. The phase difference of the light components emerging from the model is:

$$\Delta = \frac{2\pi}{\lambda} c(t_2 - t_1) = \frac{2\pi}{\lambda} d(n_2 - n_1) \quad (3.34)$$

In order to have circularly polarized light it must be $\Delta = \pi/2$; this condition makes it possible to accurately derive the thickness d of the plate, once optical properties of the material ($n_2 - n_1$) and the wavelength λ of the light in vacuum are known.

It is therefore not possible to build a Quarter-wave plate for different wavelengths or for the white light. The variation of the index of refraction is linearly related to the stresses,

$$\begin{cases} n_1 - n_0 = c_1\sigma_1 + c_2(\sigma_2 + \sigma_3) \\ n_2 - n_0 = c_1\sigma_2 + c_2(\sigma_3 + \sigma_1) \\ n_3 - n_0 = c_1\sigma_3 + c_2(\sigma_1 + \sigma_2) \end{cases} \quad (3.35)$$

where: n_0 is the index of refraction (or the refractive index) of unstressed state and n_1, n_2, n_3 are the principal refractive indices along the principal stress directions σ_1, σ_2 and σ_3 . Constants c_1, c_2, c_3 are stress optic coefficients that depend on the material.

Subtracting the previous equations member to member gives:

$$\begin{cases} n_2 - n_1 = (\sigma_1 - \sigma_2)(c_2 - c_1) \\ n_3 - n_2 = (\sigma_2 - \sigma_3)(c_2 - c_1) \\ n_1 - n_3 = (\sigma_3 - \sigma_1)(c_2 - c_1) \end{cases} \quad (3.36)$$

For a model of thickness d with its plane perpendicular to the third principal-stress direction, the first of Eq. 3.36 can be written as:

$$\Delta = \frac{2\pi}{\lambda} d(n_2 - n_1) = \frac{2\pi}{\lambda} d \cdot B(\sigma_1 - \sigma_2) \quad (3.37)$$

where: $(c_2 - c_1) = B$ is the *relative stress-optic coefficient* (brewster)

$$\left(1 \text{ Brewster} = 10^{-12} \frac{m^2}{N} = 10^{-6} \frac{m^2}{MN} = 10^{-6} \frac{1}{MPa} \right)$$

For a positive birefringence, the wave velocity related to principal stress σ_1 is greater than wave velocity related to principal stress σ_2 . So if $\sigma_1 \geq \sigma_2 \geq \sigma_3$ then $n_3 \geq n_2 \geq n_1$.

The phase difference, i.e. the relative retardation between the two vector components, is thus proportional to $(\sigma_1 - \sigma_2)$.

Let $N = \Delta/2\pi$ be the fractional phase shift and: $f_\sigma = \lambda/B$ a *photoelastic constant* of the material for the light wavelength λ . Equation (3.37) becomes:

$$\sigma_1 - \sigma_2 = f_\sigma \frac{N}{d} \quad (3.38)$$

In the elastic range are valid the Hooke's equations:

$$\begin{cases} \varepsilon_1 = \frac{1}{E}(\sigma_1 - \nu\sigma_2) \\ \varepsilon_2 = \frac{1}{E}(\sigma_2 - \nu\sigma_1) \end{cases}$$

from which the following equation is derived:

$$\varepsilon_1 - \varepsilon_2 = \frac{1 + \nu}{E}(\sigma_1 - \sigma_2) = f_\varepsilon \frac{N}{d}$$

where:

$$f_\varepsilon = \frac{1 + \nu}{E} f_\sigma \quad (3.39)$$

The inverses of the two constants: $1/f_\sigma$ and $1/f_\varepsilon$ have respectively the meaning of stress and strain sensitivities, because they express the material birefringence (fringes number), for given values of stress and strain.

3.6 Polariscopes

The instrument for measuring the birefringence at points of a loaded plane model (and consequently, the difference of the principal stress or strain with relative directions), is called a *polariscope*.

This instrument can have several forms but in its simplest arrangement it consists of a light source (white and monochromatic) and of two polarizers and two quarter-wave plates.

Two main types of polariscopes are available on the market and can be classified according to the type of light transmission:

1. The diffused light polariscope (Fig. 3.9).
2. The lens or (parallel) light polariscope.

They differ only in the possibility of having diffused or parallel light beams. The acquired images are substantially the same; the parallel rays allow some special techniques that cannot be used for the other one.

The scheme of the optical filters for a Polariscope (in parentheses the variations for lens polariscope) is the following, starting from the source along the light path:

1. A diffuse source of monochromatic light, e.g. natrium vapor lamps.
2. A second diffused source of white light alternative to the first.
3. Diffuser partially transparent plate.
 - A point source of light in the focus of a lens (for lens polariscope).
 - First convex-plane lens (for lens polariscope).
4. First Polarizer.
5. First Quarter-wave plate.
6. Plane Model with the loading apparatus.
7. Second Quarter-wave plate.
8. Second Polarizer called Analyzer.
 - Second plane-convex lens (for lens polariscope).
9. Recording Instrument (analogical or digital or video camera).

The Quarter-wave plates (*Q.O.*) can be inserted or not on the optical path.

Analyzer must be coupled with the polarizer to allow a synchronic rotation of both filters with crossed polarization axes.

The lens polariscope, with an optical system consisting of at least two lenses, allows more orthodox observations. It is useful for the definition of precise contours of the model and for the use of special techniques such as fringes multiplication. Downstream, a lens is placed after the analyzer and converges the rays in the lens focus of the recording unit or on a screen.

The diffused light polariscope (easier to build) can in all other cases replace the polariscope with parallel light, if care is taken to record the image with a medium telephoto lens to limit errors due to the inclination of the direction of the rays that converge on the lens of the recording apparatus.

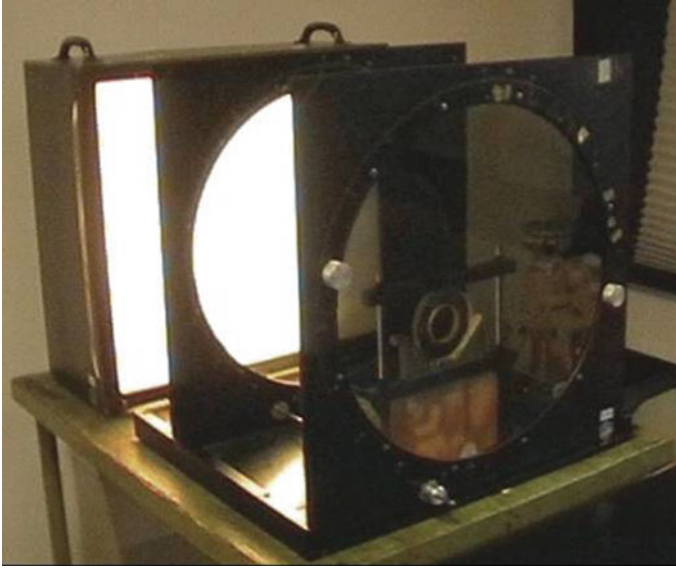


Fig. 3.9 A Diffused Light polariscope (by Tiedemann–Garmish, Germany)

3.7 Plane Polariscope

The plane polariscope consists of two plane polarizers only. It can be built very easily with two Polaroid filters and with a light source. It is used in two main configurations: with crossed or with parallel polarizing filters (Fig. 3.10).

3.7.1 Polarizer and Analyzer with Crossed Axes

Without losing generality, x and y axes can be chosen respectively in horizontal and vertical directions. The emerging light vector, in the hypothesis of a vertical polarizer and a horizontal analyzer, for Eqs. 3.17 and 3.27 is given by:

$$\mathbf{A}' = \mathbf{P}_x \mathbf{M} \mathbf{P}_y \cdot \mathbf{A}$$

then:

$$\begin{pmatrix} \mathbf{A}'_x \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \mathbf{A}_y \end{pmatrix}$$

from which:

$$\begin{pmatrix} \mathbf{A}'_x \\ 0 \end{pmatrix} = \begin{pmatrix} m_{12} \\ 0 \end{pmatrix} \mathbf{A}_y \quad (3.40)$$

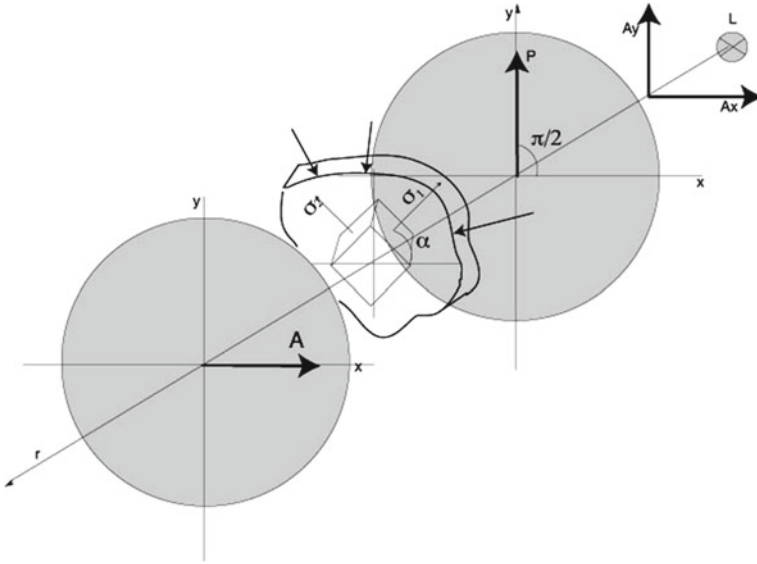


Fig. 3.10 Plane polariscope with crossed axes: source L , polarizer with polarizing axis P along y , plane model with principal stresses in the observed point, analyzer with polarizing axis A along x

with:

$$m_{12} = (1 - e^{i\Delta}) \sin \alpha \cos \alpha$$

There are two light extinction (or destructive interference) conditions:

$$e^{i\Delta} = 1 \quad \sin \alpha \cos \alpha = \sin 2\alpha = 0$$

- The first is verified when $\Delta = 0$ plus multiple of 2π i.e. $\Delta = 0 + 2k\pi$ with k integer.
- The second is verified when $2\alpha = 0$ or π .

Zeros of the function are:

$$N = \frac{\Delta}{2\pi} = 0, 1, 2, \dots, k \tag{3.41}$$

$$\alpha = 0 \quad \alpha = \frac{\pi}{2} \tag{3.42}$$

The same result is obtained from the condition of zero light intensity. For Eq. 3.32:

$$I \propto A^{*'} \cdot A' = |A_y|^2 m_{12}^* m_{12} \tag{3.43}$$

$$I \propto |A_y|^2 \sin^2 \alpha \cos^2 \alpha [2 - (e^{i\Delta} + e^{-i\Delta})]$$

Remembering the Euler's identity:

$$\cos \Delta = \frac{e^{i\Delta} + e^{-i\Delta}}{2} \quad \sin \Delta = \frac{e^{i\Delta} - e^{-i\Delta}}{2i}$$

$$I \propto 2 \sin^2 \alpha \cos^2 \alpha |A_y|^2 (1 - \cos \Delta) = 2 \sin^2 2\alpha \sin^2 \frac{\Delta}{2} = 0$$

Then $I = 0$ for:

$$\sin 2\alpha \sin \frac{\Delta}{2} = 0 \quad (3.44)$$

from which:

$$\begin{cases} \Delta = 0 + 2k\pi \\ 2\alpha = 0 \text{ or } \pi \end{cases} \quad (3.45)$$

that are the same of Eqs. 3.41 and 3.42.

3.8 Circular Polariscopes

A circular polariscopes consists of two light sources, monochromatic and white, (or a monochromatic filter to produce monochromatic light from white light), a first polarizing plate, two quarter-wave plates and a second polarizing plate. First polarizer and first quarter-wave plate (oriented at $\alpha = \pi/4$ with respect to the polarizing axis), produce circular polarized light. The second quarter-wave plate has the fast axis and slow axis inverted respect to the first one in order to inversely transform the circular polarized light into a plane polarized light. The last filter is the analyzer, Fig. 3.11.

3.8.1 Polarizer Along y Axis and Analyzer Along x Axis

The light vector emerging from the analyzer is given by, Fig. 3.11:

$$\mathbf{A}' = \mathbf{P}_x \mathbf{Q}_{-\frac{\pi}{4}} \mathbf{M} \mathbf{Q}_{\frac{\pi}{4}} \mathbf{P}_y \mathbf{A} \quad (3.46)$$

The transformation matrix is:

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad (3.47)$$

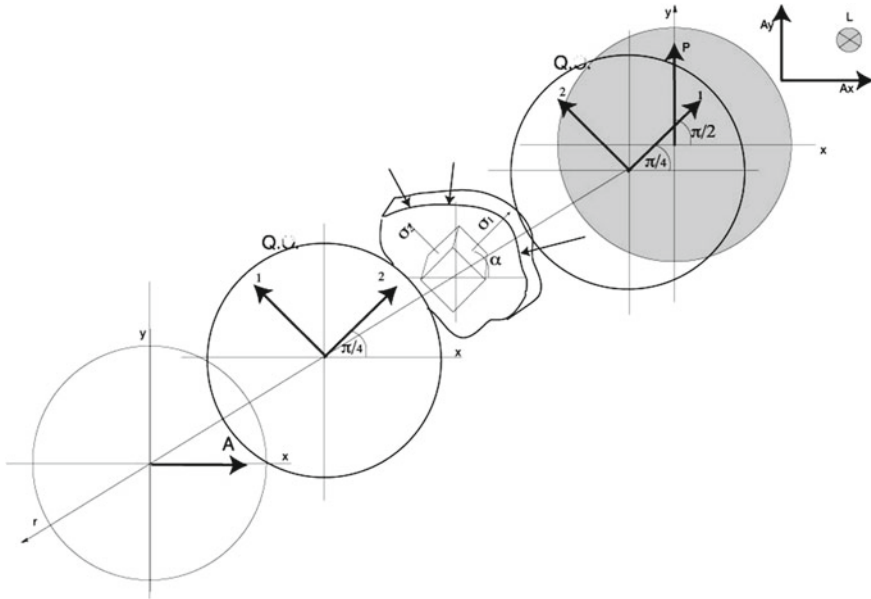


Fig. 3.11 Filters set-up in a circular polariscope: source L , polarizer P with axis along y , quarter-wave plate Q with axes 1 and 2 at $\pi/4$ and $3/4\pi$ respectively, plane model, second quarter-wave Q with axes 2 and 1 at $\pi/4$ and $3/4\pi$ respectively, analyzer A with its axis along x . Positive direction counterclockwise looking towards the light-source

from which:

$$\mathbf{A}'_x = \mathbf{A}_y(2m_{21} - im_{11} + im_{22}) \tag{3.48}$$

Substituting the expressions of m_{11}, m_{22} etc.:

$$\mathbf{A}'_x \propto \mathbf{A}_y(1 - e^{i\Delta})(\sin 2\alpha - i \cos 2\alpha) \tag{3.49}$$

There is only one light extinction condition ($1 - e^{i\Delta} = 0$), because the second term does not become zero for any value of α . Then: $e^{i\Delta} = 1$ from which Eq. 3.41 derives.

Alternatively the condition of zero light intensity is:

$$I \propto \mathbf{A}'^* \cdot \mathbf{A}' = \frac{|A_y|^2}{4}(1 - \cos \Delta) = |A_y|^2 \sin^2 \frac{\Delta}{2} \tag{3.50}$$

This equation has the same zeros as the previous one for the plane polariscope, see Eq. 3.41. The other condition given by Eq. 3.42 regarding the principal stress orientation is now absent.

3.8.2 Both Polarizer and Analyzer Along the y Axis

In this case the general expression of the transmitted light becomes:

$$\begin{aligned} \mathbf{A}' &= \mathbf{P}_y \mathbf{Q}_{-\frac{\pi}{4}} \mathbf{M} \mathbf{Q}_{\frac{\pi}{4}} \mathbf{P}_y \mathbf{A} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ A_y \end{pmatrix} \end{aligned} \quad (3.51)$$

from which:

$$\mathbf{A}'_y \propto \mathbf{A}_y (m_{11} + m_{22}) \quad (3.52)$$

$$(m_{11} + m_{22}) = (1 + e^{i\Delta}) = 0$$

verified for:

$$e^{i\Delta} = -1 \quad (3.53)$$

In the case of the analyzer along y axis, the destructive light interference gives (with k integer):

$$\frac{\Delta}{2} = \frac{2k + 1}{2} \pi \quad (3.54)$$

$$N = \frac{\Delta}{2\pi} = 0.5, \quad 1.5, \quad 2.5, \dots \quad (3.55)$$

Alternatively, the light intensity expression is, Eq. 3.31:

$$I \propto |A_y|^2 (m_{11}^* + m_{22}^*) (m_{11} + m_{22})$$

or, with the same zeros of Eq. 3.53:

$$I \propto |A_y|^2 (1 + e^{-i\Delta})(1 + e^{i\Delta}) = |A_y|^2 \frac{1 + \cos \Delta}{2} = |A_y|^2 \cos^2 \frac{\Delta}{2} \quad (3.56)$$

3.9 Isochromatics

As it was previously shown with Eq. 3.38, the difference of the principal stresses in a point of a plane model is directly proportional to the photoelastic constant, to the parameter N and inversely to the thickness d at that point.

$$\sigma_1 - \sigma_2 = f_\sigma \frac{N}{d} \quad (3.57)$$

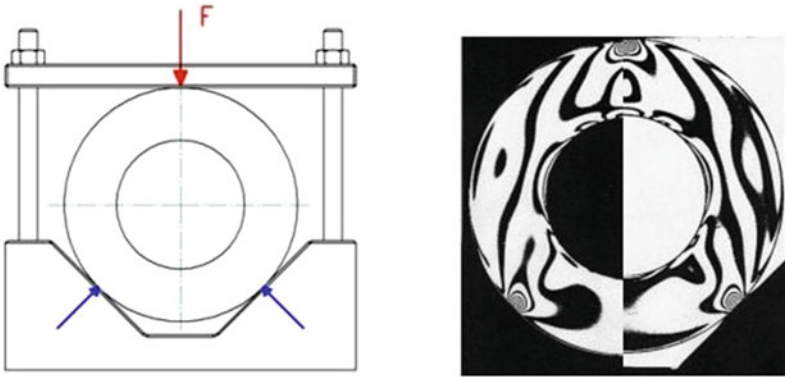


Fig. 3.12 Isochromatics on a black and bright background of an annular disc loaded at three points

If the observation is carried out in monochromatic light with wavelength λ , the loci of points of light extinction with crossed filters (or the maximum transmission with parallel filters), are interference fringes that are called *isochromatics*. So such fringes appear black on a black background with crossed filters and bright on a bright background with parallel filters (Fig. 3.12). Details of the contact points are shown in Fig. 3.13.

Isochromatics join the points having the same value of the difference $\sigma_1 - \sigma_2$, namely the same diameter as the Mohr's circle, i.e. the same maximum shear stress;

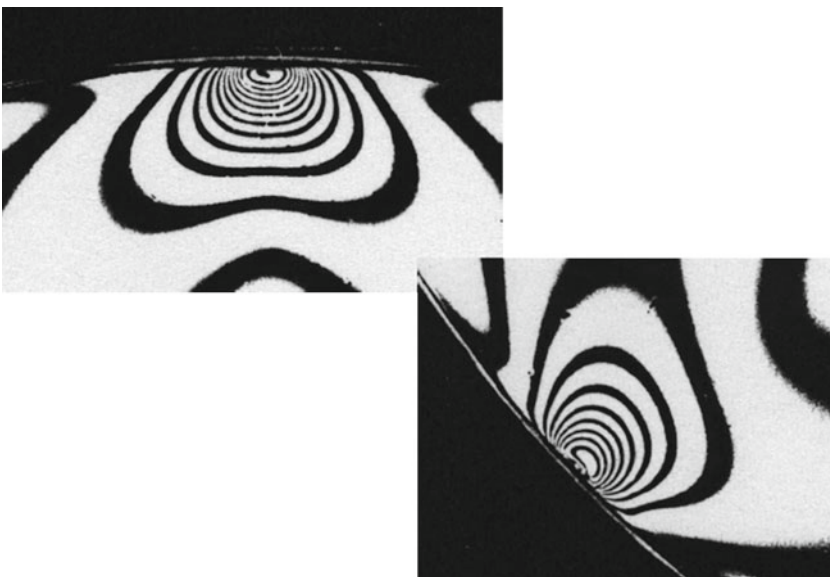


Fig. 3.13 Details of isochromatics in contact points without (*up*) and with a light friction (*bottom*)

Table 3.1 $N\lambda$ values up to the first order with crossed polarizers in white light, modified from [13]

$N\lambda$ (nm)	Absorbed	Transmitted
0	All	Black
40		Iron gray
57		Lavender gray
158		Gray blue
218		Gray
234		Green white
259		White
267		Yellow white
275		Pale straw yellow
281		Straw yellow
308		Bright yellow
390	Indigo-violet	Brilliant yellow (corr. to $N = 1$ with sodium light)
430		Brown yellow
505	Blue-green	Orange
536	Green	Red
551		Deep red
565		Purple
576*	Brilliant yellow	Indigo-violet (passage from warm to cold colors)
664	Orange	Sky blue
728		Green blue
747	Red	Green
800		Bright green
826		Yellow-green

*Correspondent to order 1 with the use of Sodium vapor light

in other words they are the contour lines of the difference of the principal N is called *Isochromatic order*.

If observation is carried out in white light, isochromatics appear colored by the complementary colors of the extinguished ones,⁵ whose wavelength λ is linked to the principal stresses difference by the relationship, Eq. 3.37:

$$B \cdot (\sigma_1 - \sigma_2)d = N\lambda \quad (3.58)$$

with N given respectively by Eqs. 3.41 and 3.55 for crossed and parallel polarizers. In this way, each value of $d(n_2 - n_1)$ i.e. $(\sigma_1 - \sigma_2)$ is distinguishable for a characteristic color of transmitted light, Table 3.1.

⁵ The colors of all wavelengths but the extinct one, are transmitted and perceived as a single color, called the complementary color of the extinguished one.

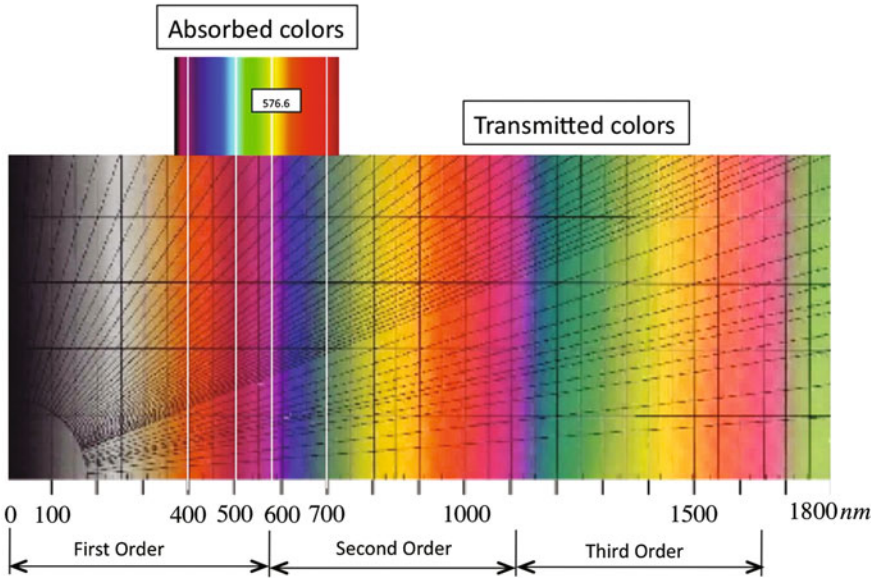


Fig. 3.14 Absorbed and transmitted color: modified from [16]

Let us now continuously increase the difference ($\sigma_1 - \sigma_2$), starting from zero value. If the principal stresses difference is zero, the transmitted light intensity is zero and a black fringe appears. The light of all wavelengths is absorbed and order $N = 0$ occurs.

For higher stress differences, (keeping $N = 1$), extinction happens for wavelengths smaller than the minimum of the visible spectrum (violet) and the transmitted light varies from black, through shades of gray, up to white at about 260 nm. The first colors clearly defined appear due to the extinction of a wavelength of approximately 390 nm (violet) ($N\lambda = 1 \cdot 390 \text{ nm}$), and the transmitted light color, complementary to the absorbed one is yellow. The extinction occurs for the following wavelengths of the visible spectrum in this order: blue, green, yellow, orange and red. The colors are then transmitted respectively in the sequence: yellow, orange, red, purple, blue, green, Fig. 3.14.⁶

At about 577 nm extinction of the first order of colors $N = 1$ is complete. If the difference ($\sigma_1 - \sigma_2$) increases, extinctions begin for twice the wavelengths of the visible spectrum ($N = 2$) and the colors are extinguished a second time. If the difference becomes higher, the number of wavelengths cancelled at the same time tends to increase. The second order is complete at about 1100 nm. From the third order the transmitted colors are the sum of several complementary colors and appear faded, tending to pink, pale green, and finally, to white. Theoretically, the estimation

⁶ Taken from Prof. Stephen A. Nelson, Department of Earth and Environmental Sciences. EENS 2110, Mineralogy, Tulane University: Interference Phenomena, Compensation, and Optic Sign.

of the color of the transmitted light could offer a way to calculate the difference of the principal stresses. Since the isochromatic of integer order corresponds to a precise multiple of the wavelength of the monochromatic light used, the estimation of an intermediate color in white light would make it possible to evaluate the fractional order of isochromatic. However, estimation is uncertain and it is preferable to use white light only for identifying the isochromatic of zero order, which appears always black.

Conversely, the accurate counting of of isochromatic orders is done in monochromatic light, because they appear sharp up to very high (10 and more) orders. If sodium vapor lamps are used ($\lambda = 576.6 \text{ nm}$), the first order corresponds to the indigo-violet color observed in white light.

3.10 Isoclines

It is possible, with the use of plane polarized light, to show the condition of light extinction expressed by the relationship:

$$2\alpha = n\pi$$

This happens for the points of the model that have the principal stress directions coinciding with polarization directions of polarizer and analyzer with crossed axes.

The locus of these points of equal inclination of the principal directions are continuous curves called *isoclines*.

With the use of circular polarized light, isoclines are deleted and only the isochromatics appear. It is not possible to operate the inverse, i.e., to eliminate the isochromatics and keep only the isoclines; nevertheless, it is possible to distinguish these from the first ones following these methods:

- Rotate polarizer and analyzer maintaining their relative directions at 90° . The isocline changes position on the model surface and, during a complete rotation of 90° , cover all the points of the model, while isochromatics remain unchanged.
- The model is observed in white light. The isochromatics are colored, while the isoclines appear as dark lines or bands.
- Reduce the applied loads gradually. The isoclines, that depend only on the principal directions and not on the stress values, remain unaltered, while the isochromatics tend to disappear.

It should be noted that, generally, the isoclines have a band shape with a certain width. The midline is the isocline value for the assigned inclination of the polarizer axes.

If the isocline is recorded for a number of discrete values of the inclination of the polarizing filters, a family of isoclines is generated, each of which can be marked by the value of the inclination from which it derives.

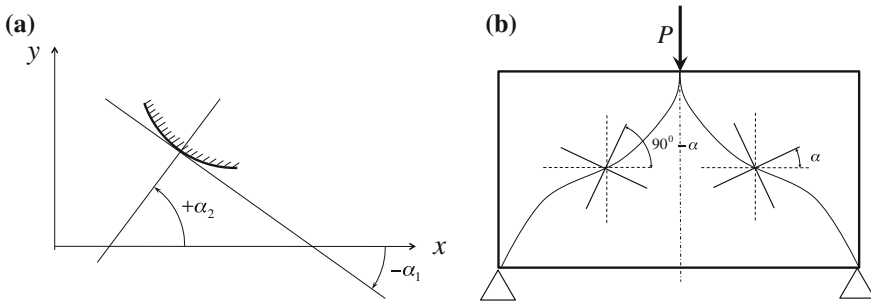


Fig. 3.15 Isoclines properties: Isoclines at a free boundary (a) and isoclines at symmetric points (b)

The isoclines have special attributes resulting from their definition, Fig. 3.15:

1. The isocline parameter depends on the initial position of the model with respect to the reference x , y identified by the initial directions of the analyzer and polarizer axes.
2. The isocline parameter depends on the rotation direction of the crossed filters. The same isocline has values complementary to 90° if it is derived by clockwise or counterclockwise rotation. Consequently, symmetrical isoclines have complementary to 90° values.
3. The isocline at a point on a free boundary has, by definition, the value of the angle formed by the tangent (or by the normal) to the border with the x axis, depending on the direction of the crossed filters rotation.
4. The points of a straight free edge belong to the same isocline.
5. In the case of geometrical and physical (forces distribution) symmetry, the symmetry axis coincides with an isocline.
6. All isoclines pass through the points that have equal principal stresses.

3.10.1 Isotropic and Singular Points

If the difference $\sigma_1 - \sigma_2$ is equal to zero at a certain point, that point has optical isotropic behavior and is called *isotropic*.

Moreover, if $\sigma_1 = \sigma_2 = 0$ the point is without any stress and is called *singular*.

If an isotropic point is on a free edge of the model, it is necessarily singular. This point indicates, in general, the change of sign of the principal stress tangent to the boundary.

By rotating the filters of the polariscope in a given direction it may happen that the isocline at an isotropic point rotates in the same or in the opposite direction of the filters.

Points of the first type are called *positive* and the other *negative*. A *singular* point on the free edge of the model is necessarily negative.

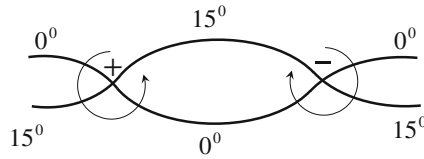


Fig. 3.16 Properties of isotropic points: positive when the isoclines sequence rotates in the same direction as the polariscope axes, negative when the isocline sequence is in the opposite direction

The isotropic contiguous points are alternately positive and negative, since the sequence of the isocline at one point is necessarily reverse at the next point, Fig. 3.16.

The *order* of an isotropic point is defined by the number of isoclines of equal value that pass through it. If only one isocline is present at the point with a single polarizer direction (in the interval $0-90^\circ$), the isotropic point is of the *first order*.

The isoclines do not coincide only at the isotropic points; coincidence happens also in application points of concentrated loads, i.e. singularities for stress tending to infinity. In some cases, instead of single points there is an isotropic region.

3.10.2 Determination of Fractional Fringe Orders

One of the problems of the information produced by the isochromatics curves is the necessity of interpolation for stress calculation when the point does not lie exactly on the isochromatic fringes and phase shift does not correspond exactly to an integer multiple or half-integer order.

Small phase shifts (or fractional orders) can be accurately measured by instruments called *compensators* or by a special set-up of optical filters of the polariscope, according e.g. to Tardy or Senarmont methods.

The first way to determine the fractional orders of isochromatics and to identify the principal stress directions is by the use of a simple optical device that introduces, on the optical light path in the polariscope, a known value of phase shift that equalizes the unknown phase shift in the point of the model under examination. The device therefore compensates the unknown value with a known one (which is the reason for its name), with the same philosophy of the two pans balance.

3.10.2.1 Compensator Made from a Wedge Pre-compressed Bar

The concept underlying these tools can be clarified for the simplest type of compensator. It is built with a small bar of photoelastic material, uniaxially loaded in compression along its axis.

The procedure for making a compensator with adjustable compressive state is the following:

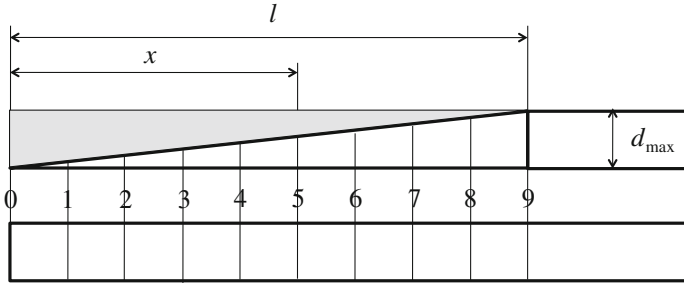


Fig. 3.17 Compensator made by a wedge bar with frozen compression stress

A compression uniaxial stress state is *frozen* in a small parallelepiped bar of photoelastic material, Fig. 3.17.

The bar is cut in a wedge shape as shown in the same figure. According to the following formula derived from the general equation of photoelasticity, Eq. 3.38:

$$N = \frac{\sigma}{f_{\sigma}} \frac{d_{max}}{l} x \tag{3.59}$$

the isochromatic order is proportional to x , i.e. the distance of the generic point from the wedge edge. Isochromatics are thus linearly distributed along the length of the bar, at an equal distance from one another. The compensator is finally completed by bonding this wedge-bar to another with the same shape, but stress-free, in order to avoid spurious refraction effects in the light transmission. Along its length orders of isochromatic are marked like a ruler, Fig. 3.17.

Let 1 and 2 be the principal stress directions at a point of a (loaded) model in the polariscope. The longitudinal direction of the compensator is placed, in turn, along the principal directions over the point. The effect of birefringence due to the overlapped state of the uniaxial stress in the beam (that is known and variable with continuity), and the unknown birefringence at the point of the model, Fig. 3.18 is observed.

If the unknown stress state at the point of the model is $(\sigma_1 - \sigma_2)$, two cases are possible:

1. If the direction of the compensator coincides with the principal direction 1, the overall retardation (or the overall phase shift) decreases: augmenting the compression level of the compensator, decreases the order of the optical system consisting of the model and compensator, eventually reaching zero, (shown by a black isochromatic of zero order). The unknown phase shift (fractional order) in the model is equal and opposite to that produced in the bar loaded with a known compressive stress.
2. If, vice-versa, the bar axis is superimposed on the direction 2, the phase shift in the composed optical system increases, i.e. the isochromatic assumes different colors and the condition of light extinction (black isochromatic) is never reached

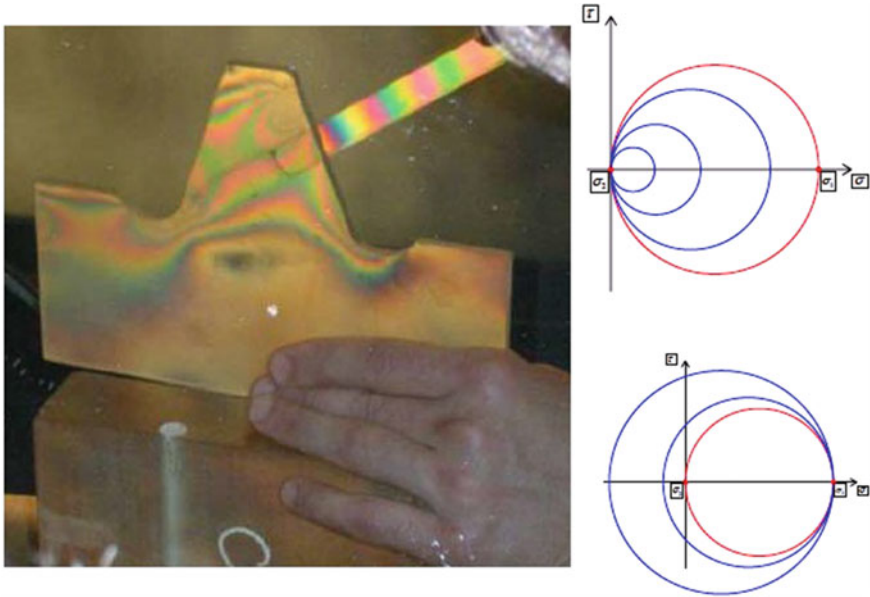


Fig. 3.18 Illustration of the use of compensator: a known optical effect is superimposed to the unknown one in the model, with identical effect of an added compression stress

for whatever value of compressive stress in the bar. In this case, the measurement of the order is impossible.

Through this procedure the principal stress directions are identified and the order is measured.

If the analysis is done at points on free boundaries, the technique gives the only principal stress not equal to zero, tangent to the boundary; from Eq. 3.38 thus the following derives:

$$\sigma_T = f_\sigma \frac{N}{d} \quad (3.60)$$

and the phase shift measured by the compensator gives the unknown stress.

The isochromatic orders can be read directly on the compensator, Fig. 3.18. This order is equal and opposite to the unknown fractional fringe order at that point.

There are other methods, in addition to the one described above that indicate which of the principal directions is associated with the higher stress:

1. By examination of the isoclines and isostatics.
2. By a concentrated load at the free boundary, Fig. 3.19.

This method consists in the observation of the shape of the isochromatic on the free boundary due to the effect of a localized pressure produced by a pointed object in a direction perpendicular to the surface. If the tangent direction is in a tensile state, the pressure $\sigma_2 = -p$ produces a phase shift between principal stresses such that:

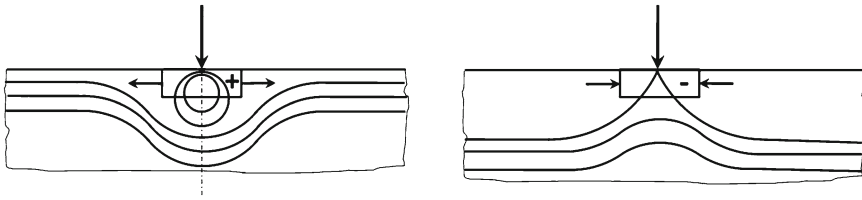


Fig. 3.19 Method of localized pressure to determine the stress sign at a free surface of a loaded plane model

$$(\sigma_1 - \sigma_2) > \sigma_1$$

So the order increases, new isochromatics enter the model and isochromatics of a lower order are pushed inside the model. In the opposite case of a tangent compressive state, the following holds:

$$(\sigma_1 - \sigma_2) < \sigma_1$$

and the isochromatics with lower order tend to move towards the boundary and the order at the edge decreases.

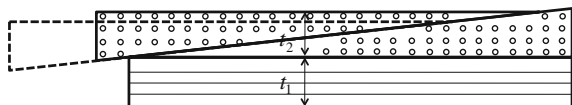
3.10.2.2 Babinet-Soleil Compensator

The Babinet-Soleil compensator is a variant but also an improvement of the previous one, since it allows the use of positive or negative retardation (phase shift) to add to the unknown retardation at the point of the model (Fig. 3.20).

It is made from a small parallelepiped of quartz of uniform thickness t plus two quartz wedges that have the optical axes perpendicular to the first. The thickness of the double-wedge is variable with the lateral position of one of the wedges which is mobile. In this way, the total phase shift introduced by the compensator in an optical path is proportional to the difference between the thicknesses of the bar of constant dimension and the bar made from the two wedges. In the system consisting of the parallelepiped bar and by two wedges, when $t_1 = t_2$ the overall retardation is zero; for $t_1 \neq t_2$ retardations are positive or negative.

The compensator works in this way: having chosen the point in the model and identified the principal directions, the compensator must be aligned with one of them and the thickness adjusted to extinguish the phase shift at that point. The position

Fig. 3.20 Babinet-Soleil compensator



of the lateral micrometer screw that moves the wedge is calibrated in order to be proportional to the isochromatic unknown order at the point.

3.10.3 Measurement of Fractional Orders by Means of Quarter-Wave Plates

Fractional isochromatic orders are determined more frequently with other methods that use only the normal equipment of the polariscope. Theoretically, through these methods a resolution of one-hundredth of an order can be reached.

3.10.3.1 Tardy Method

In the *Tardy* method the analyzer itself acts as a compensator. Having chosen the point for the examination, the axis of the polarizer is aligned along the direction of σ_2 and the other optical elements of the circular polariscope are arranged to give a dark field, Fig. 3.21. In general, as the first operation, the polariscope axes x and y are rotated up to coincide with the principal stress directions at the point. For clarity, the direction of σ_1 is assumed to be coincident with the axis x in a horizontal direction.

The light vector emerging from the analyzer when the analyzer with its vector A is rotated counterclockwise by ϑ with respect to x axis, (that coincides with the initial direction of the vector A_0), is given by:

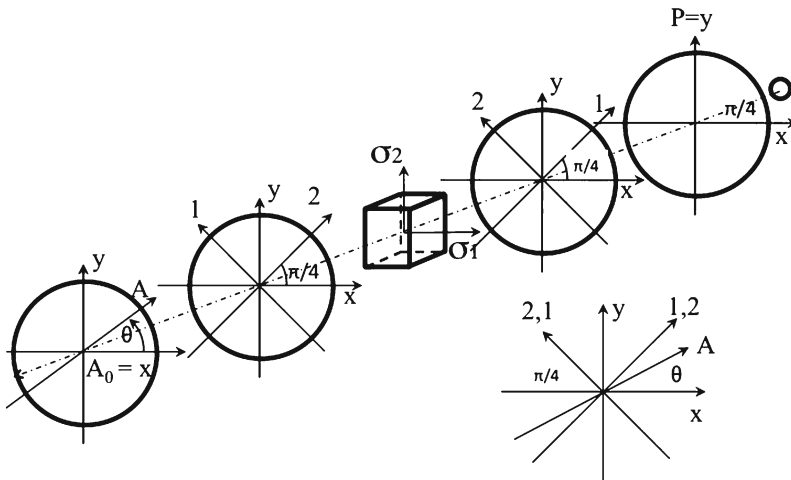


Fig. 3.21 Measurement of fractional order by Tardy method

$$\mathbf{A}' = \mathbf{P}_\vartheta \mathbf{Q}_{-\frac{\pi}{4}} \mathbf{M}_{\alpha=0} \mathbf{Q}_{\frac{\pi}{4}} \mathbf{P}_y \mathbf{A} \quad (3.61)$$

The transformation matrix, a product of matrices of Eq. 3.61, taking into account Eqs. 3.16 and 3.26, is:

$$\begin{aligned} & \begin{pmatrix} \cos^2 \vartheta & \sin \vartheta \cos \vartheta \\ \sin \vartheta \cos \vartheta & \sin^2 \vartheta \end{pmatrix} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{i\Delta} \end{pmatrix} \begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\ & = \begin{pmatrix} \sin \vartheta \cos \vartheta - i \cos^2 \vartheta + i e^{i\Delta} \cos^2 \vartheta + e^{i\Delta} \sin \vartheta \cos \vartheta & \\ \sin^2 \vartheta - i \sin \vartheta \cos \vartheta + i e^{i\Delta} \sin \vartheta \cos \vartheta + e^{i\Delta} \sin^2 \vartheta & \end{pmatrix} \end{aligned}$$

For Eq. 3.32, unless the term $\frac{1}{4}|A_y|^2$, the intensity expression is:

$$\begin{aligned} I \propto & \sin^2 \vartheta \cos^2 \vartheta (1 + e^{-i\Delta})(1 + e^{i\Delta}) + \cos^4 \vartheta (1 - e^{-i\Delta})(1 - e^{i\Delta}) \\ & + \sin^2 \vartheta \cos^2 \vartheta (1 - e^{-i\Delta})(1 - e^{i\Delta}) + \sin^4 \vartheta (1 + e^{-i\Delta})(1 + e^{i\Delta}) \\ & + i \sin \vartheta \cos^3 \vartheta (1 + e^{i\Delta})(1 - e^{-i\Delta}) - i \sin \vartheta \cos^3 \vartheta (1 - e^{i\Delta})(1 + e^{i\Delta}) \\ & - i \sin^3 \vartheta \cos \vartheta (1 - e^{i\Delta})(1 + e^{-i\Delta}) + i \sin^3 \vartheta \cos \vartheta (1 + e^{i\Delta})(1 - e^{-i\Delta}) \end{aligned}$$

Taking into account the *Euler* relationships, is:

$$I \propto \frac{1}{2}|A_y|^2 [1 - \cos(\Delta - 2\vartheta)]$$

or:

$$I \propto |A_y|^2 \sin^2 \left(\frac{\Delta}{2} - \vartheta \right) \quad (3.62)$$

The condition of light intensity extinction is given by:

$$\frac{\Delta}{2} - \vartheta = n\pi$$

with n integer. or:

$$N = \frac{\Delta}{2\pi} = n + \frac{\vartheta}{\pi} \quad (3.63)$$

Rotating the analyzer in the direction of $\vartheta < 0$ (clockwise), the condition for light extinction is the following:

$$N = \frac{\Delta}{2\pi} = n - \frac{\vartheta}{\pi} \quad (3.64)$$

In order to avoid errors it is advisable to use a crossed polarizer and analyzer to detect the orders of the isochromatic adjacent to the point under examination. The procedure is the following:

- Remove the quarter-wave plates and rotate the polarizer and analyzer together until reaching the condition of light extinction. With this operation the isocline is brought to the point.
- Return the quarter-wavelength plates in their relative position at $\pi/4^\circ$ with respect to the new direction of the polarizers.
- Rotate the analyzer of an angle ϑ to obtain the light extinction: turning to anti-clockwise or clockwise, bring the closest isochromatic of integer order to the point.

If the order of this isochromatic is of order n_{i-1} the extinction condition is reached for a fractional order:

$$n_{i-1} + \frac{\vartheta_1}{\pi} \tag{3.65}$$

If the order of this isochromatic is n_i , the extinction condition is reached for a fractional order:

$$n_i - \frac{\vartheta_2}{\pi} \tag{3.66}$$

3.10.3.2 Senarmont Method

Also in this case, without loss of generality, the principal stress directions at the point are assumed respectively along the horizontal and vertical axes (Fig. 3.22).

The method needs the following steps:

- The plane polariscope is rotated to bring light extinction to the point, i.e. to bring the isocline in it.
- The polariscope filters are rotated to bring the polarizer and analyzer axes at 45° with respect to the principal directions.

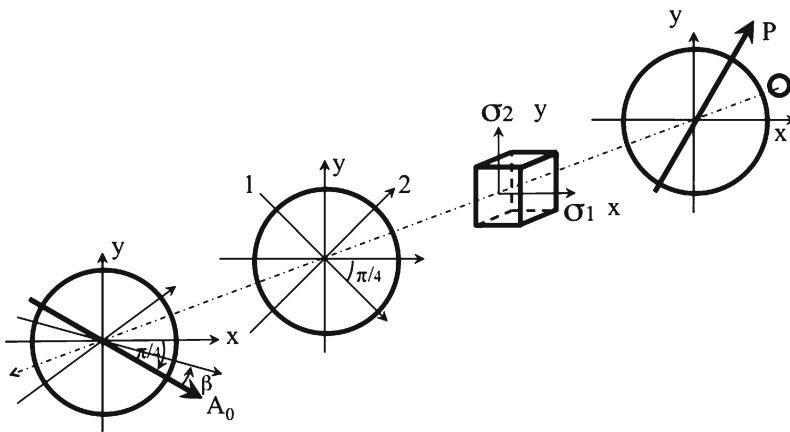


Fig. 3.22 Measurement of fractional order by Senarmont method

- The second quarter-wave plate downstream of the model must be fitted with one of the polarizing axes parallel to the analyzer axis.
- The analyzer from the position of the A_0 must be rotated by an angle $+\beta$ i.e. with respect to the x axis of an angle $\vartheta' = -\frac{\pi}{4} + \beta$ to bring an isochromatic of integer order to the point.

With these operations, the emergent light vector is given by:

$$\mathbf{A}' = \mathbf{P}_{\beta-\pi/4} \mathbf{Q}_{-\pi/4} \mathbf{M}_{\alpha=0} \mathbf{P}_{\pi/4} \mathbf{A} \quad (3.67)$$

The transformation matrix is:

$$\begin{pmatrix} \cos^2 \vartheta' & \sin \vartheta' \cos \vartheta' \\ \sin \vartheta' \cos \vartheta' & \sin^2 \vartheta' \end{pmatrix} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} e^{i\Delta} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

After some mathematical manipulations, the following expression is obtained for the light intensity:

$$I \propto |A_y|^2 \sin^2 \left(\frac{\Delta}{2} - \beta \right) \quad (3.68)$$

that is the same found by the Tardy method.

Also in this case the extinction of light is achieved for the conditions Eqs. 3.65 and 3.66.

3.10.3.3 Method of Isochromatics Multiplication

With a proper use of two semi-reflective mirrors, the number of fringes that can be observed in a photoelastic model can be multiplied [17]. This method has two advantages compared to the compensation methods [18]: it operates instantaneously and simultaneously at all points of the model.

Two semi-reflective mirrors are placed in a parallel light polariscope on both sides of the model, according to the scheme of Fig. 3.23. The second mirror is slightly tilted with respect to the first. The effect of the inclination on the beams is shown in the same figure. It is clear that each light beam emerges from the system of mirrors with an angle that depends on the number of times that the beam has passed back and forth through the model. For e.g. rays 1, 3, 5, that have passed the model a number of times identical to their numbers, emerge with angles 0 , 2Φ , 4Φ .

Although beams do not pass strictly through the same point, the paths are much closer than it appears in the figure where the angle is amplified for clarity. The path length on which the photoelastic effect is measured as the sum of partial photoelastic effects depends on Φ , on the number of beams and on the distance between the mirrors.

In practical terms one can operate a multiplication of isochromatics from 5 to 7 times without introducing errors not acceptable due to the different paths of light in the model thickness (Φ , in practice, is approximately $1/200$ rad).

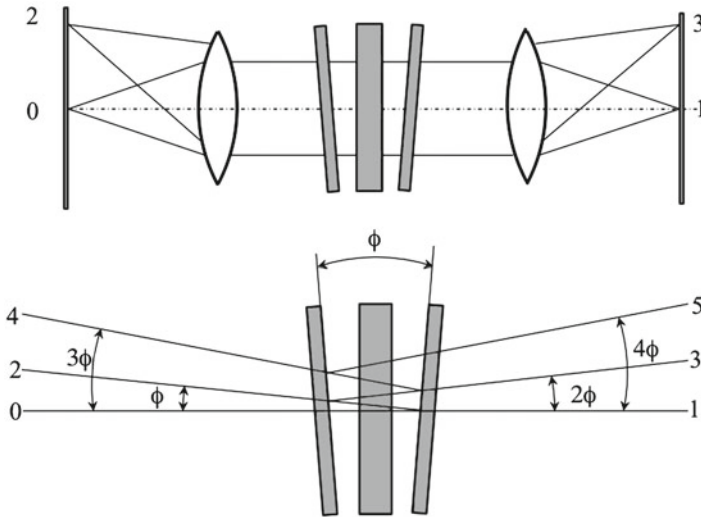


Fig. 3.23 Method of fringes multiplication

As the different beams are inclined at different angles to the axis of the polariscope, many images are formed. Each of these rays can be isolated by placing the eye or the recording device on the image of interest. It proves useful to introduce a diaphragm to delete all images, but the observed one.

Suppose, e.g., that we are recording images respectively in bright field and dark-field for beam 1.

The orders of the isochromatics are:

0, $1/2$, 1, $3/2$, 2, $5/2$, ... ,

The images obtained with ray 3 thus give orders:

0, $1/6$, $1/3$, $1/2$, $2/2$, $5/6$, 1.

The result is a densification of the isochromatics that allows a very accurate order estimation at every point of the model.

3.10.4 Drawing Stress Trajectories from Isoclines

The two families 1 and 2 of *isostatic* curves (or isostatics) can be plotted graphically once the full family of isoclines is known, drawn for a number of inclination angles of the polarizer and analyzer with respect to a reference coordinate system.

The simplest method consists in tracing many small crosses on the isocline curves oriented in the same way as the isocline angle. These crosses give the directions of both isostatics families at every point. Isostatics must be completed by joining the arms of the crosses.

A more practical method could be the following:

- All the isoclines, e.g. every 15° , are drawn, having defined a positive rotation direction of the filters, e.g. counterclockwise (looking towards the model from the analyzer).
- After fixing an angular reference, a segment of an isostatic, starting from any point of an isocline, must be drawn with an inclination corresponding to its angular value. This straight line is extended up to about the half distance from the contiguous isocline.
- Starting from the end of this segment, a new straight-line segment is drawn with a slope corresponding to this second isocline, again extending it up to about half way between this isocline and the third one. The process is repeated up to the last isocline, Fig. 3.24a.

The isostatics have some properties:

1. Curves of the family σ_1 intersect curves of σ_2 at 90° .
2. Close to the isotropic points, isostatics have a characteristic pattern shown in Fig. 3.24b, c.

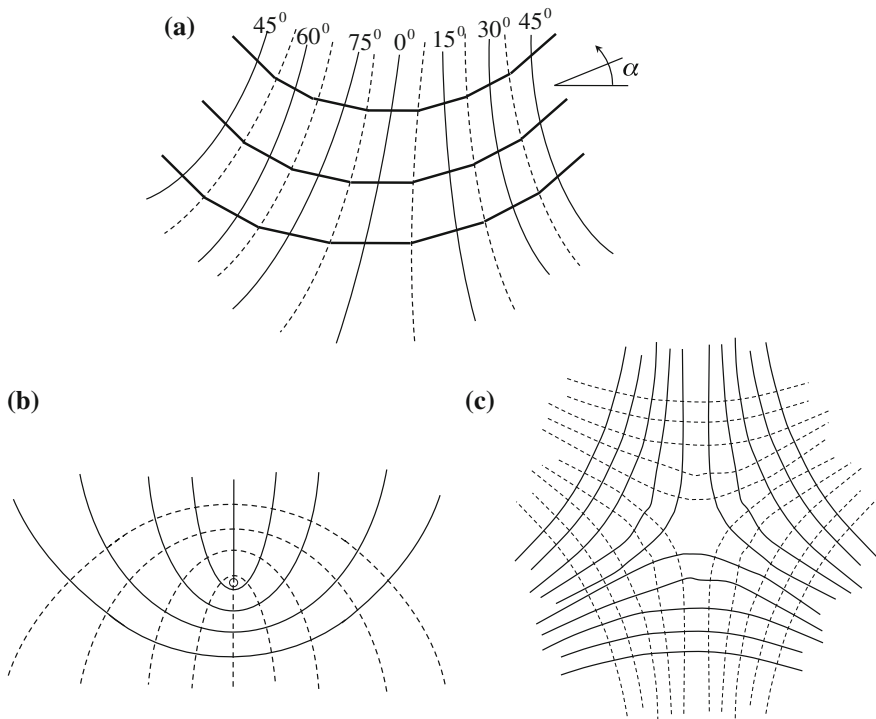


Fig. 3.24 Plotting isostatics from isoclines: **a** Typical isotropic points: **b** positive, **c** negative

3. At the straight edges of the free model boundary or at the boundary loaded with distributed perpendicular forces, the two families are respectively perpendicular and tangent to the boundary.
4. Density and bending of isostatics are related to the stress concentration in that area. Some cases will be shown in subsequent chapters.
5. Symmetry axes belong to an isostatic (in addition to an isocline);
6. Isostatics concentrate and converge in areas of loads application.
7. Principal stresses have a maximum or a minimum in the straight portion or in inflection points of the isostatics. These properties are summarized in the case-study in Fig. 3.25.

In fact, the stress equilibrium equations can be written along isostatics, according to *Lamé-Maxwell* equations in this way:

$$\begin{aligned} \frac{\partial \sigma_1}{\partial S_1} + \frac{\sigma_1 - \sigma_2}{\rho_2} &= 0 \\ \frac{\partial \sigma_2}{\partial S_2} + \frac{\sigma_1 - \sigma_2}{\rho_1} &= 0 \end{aligned} \tag{3.69}$$

Stresses σ_1 and σ_2 have a maximum or a minimum value respectively when S_2 and S_1 are rectilinear or with an inflection point (ρ_2 and respectively ρ_1 equal to ∞).

- $\frac{\partial \sigma_1}{\partial S_1} = 0$ for $\frac{1}{\rho_2} = 0$ S_2 has a straight portion or an inflection point.
- $\frac{\partial \sigma_2}{\partial S_2} = 0$ for $\frac{1}{\rho_1} = 0$ S_1 has a straight portion or an inflection point.

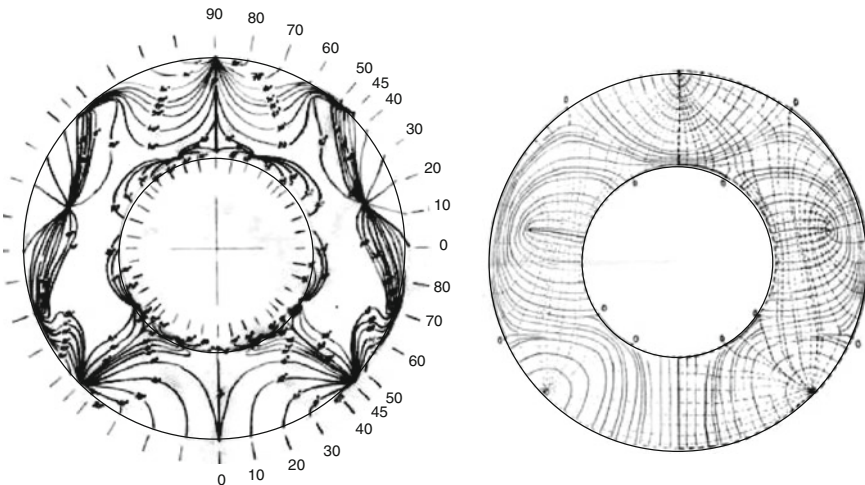


Fig. 3.25 Isoclines and deduced isostatics in an annular disk loaded at three points, Fig. 3.12

3.11 Principal Stresses Separation

The fundamental equation of photoelasticity allows direct determination of single principal stress at a point on a free boundary if the fringe order, the thickness and the photoelastic constant f_σ are known.

At internal points, only the difference of the principal stresses and their directions can be determined, i.e. stresses are not separately known.

Let us examine how it is possible to separate the principal stresses at internal points of a plane model. Some methods require further experimental data while others, based on the integration of equilibrium equations, complete the experimental data with theoretical assumptions. Among them the method of *shear stress difference* can be applied not only to the 2-D cases but also to three-dimensional ones.

3.11.1 Shear Stresses and Normal Stresses Difference Determination

Photoelasticity allows the complete knowledge of the values and signs of shear stresses and of the normal stresses difference at any point of a model, since these data derive from the knowledge of isochromatics and isoclines.

It is worth observing the Mohr's circle does not give the physical sign of τ_{xy} but the different sign describes only the fact that on two perpendicular edges the shear stress converges or diverges:

$$\tau_{xy} = \pm \frac{1}{2}(\sigma_1 - \sigma_2) \sin 2\alpha \quad (3.70)$$

In this equation α is the isocline parameter at the point.

In a similar way the stresses difference is derived:

$$\sigma_x - \sigma_y = \pm(\sigma_1 - \sigma_2) \cos 2\alpha \quad (3.71)$$

Regarding signs for τ_{xy} and for the difference $\sigma_x - \sigma_y$, the following considerations must be made.

3.11.2 Determination of Stress sign

The (conventional) signs of τ_{xy} and of the difference $\sigma_x - \sigma_y$ may be assigned by equilibrium considerations, see Fig. 3.26, once the principal stress directions are separately identified.

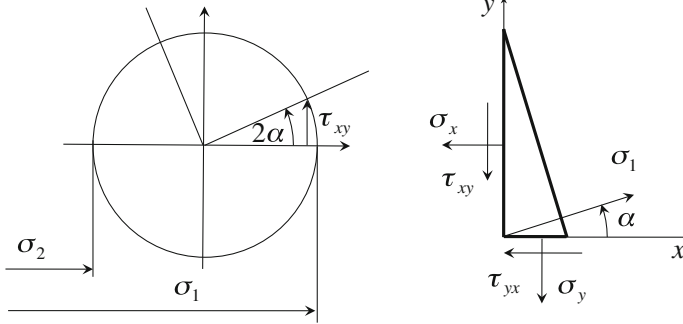


Fig. 3.26 Shear stress signs from equilibrium conditions

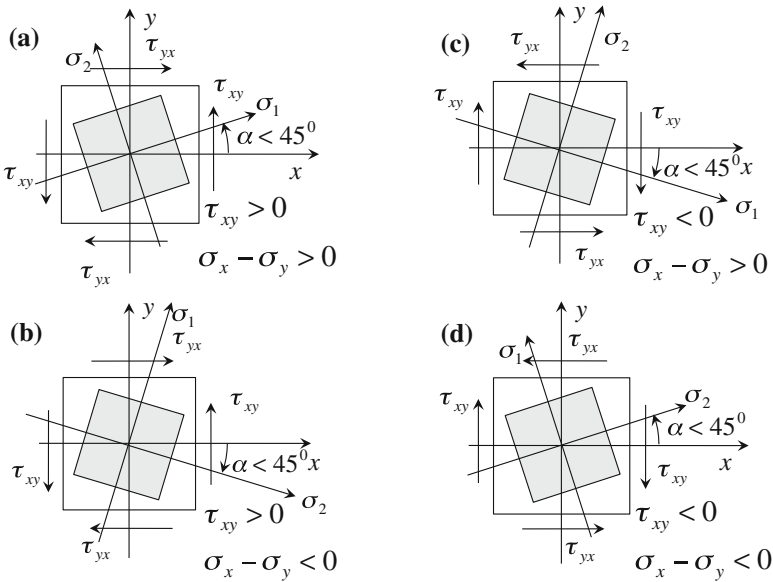


Fig. 3.27 How to obtain signs of shear stresses and of stress difference by compatibility conditions

Signs are also derived by an intuitive compatibility observation, Fig. 3.27:

- If the positive direction of σ_1 is in the first quadrant, Fig. 3.27a, τ_{xy} is positive; if also the angle between the direction of 1 and x is less than 45° , then $\sigma_x > \sigma_y$ and Eq. 3.71 must be assumed to be positive. If, vice-versa, the angle between the direction of 1 and x is greater than 45° , Fig. 3.27b, $\sigma_x < \sigma_y$ and Eq. 3.71 must be assumed to be negative.
- If the positive direction of σ_2 is in the first quadrant, τ_{xy} is negative; if also the angle between the direction of 2 and x is greater than 45° Fig. 3.27c, then $\sigma_x > \sigma_y$ and Eq. (3.71) must be assumed to be positive; if, vice-versa, the angle between

the direction of 1 and x is less than 45° , Fig. 3.27d, then $\sigma_x < \sigma_y$ and Eq. 3.71 must be assumed to be negative.

3.11.3 Integration of Equilibrium Equations

Given a plane model, after fixing a Cartesian coordinate system with the origin at a point of the free boundary of the model, the differential equilibrium equations must be integrated in order to determine the stress state at all points along a segment of the x axis. The equilibrium equations are:

$$\begin{cases} \frac{\partial \sigma_x}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} = 0 \\ \frac{\partial \sigma_y}{\partial y} + \frac{\partial \tau_{xy}}{\partial x} = 0 \end{cases}$$

If the first member is integrated with respect to x , between values 0 and i , the following relationship is derived:

$$\int_0^i \frac{\partial \sigma_x}{\partial x} dx = - \int_0^i \frac{\partial \tau_{xy}}{\partial y} dx \quad (3.72)$$

i.e.:

$$\sigma_x = \sigma_{x0} - \int_0^i \frac{\partial \tau_{xy}}{\partial y} dx \quad (3.73)$$

Once the isochromatic pattern and related isoclines (assuming x as the initial analyzer direction) are plotted, the determination of σ_x at point 0 is obtained by equilibrium along x direction of the infinitesimal element shown in Fig. 3.28.

The τ_{xy} sign is determined by equilibrium in y direction. In the case of Fig. 3.28, e.g., σ_{x0} and τ_{xy} are both positive according to the most common sign convention. The integration can be carried out for finite differences Δx , Δy .

The following steps are obtained:

1. The integral approximated by finite differences:

$$\int_0^i \frac{\partial \tau_{xy}}{\partial y} dx \simeq \sum_0^i \frac{\Delta \tau_{xy}}{\Delta y} \Delta x \quad (3.74)$$

2. The initial value given by equation, Fig. 3.28:

$$\sigma_{x0} = \sigma_0 \cos^2 \alpha \quad (3.75)$$

3. N and α along $A - A'$ and $B - B'$ at a distance $\pm \Delta y/2$ from x .

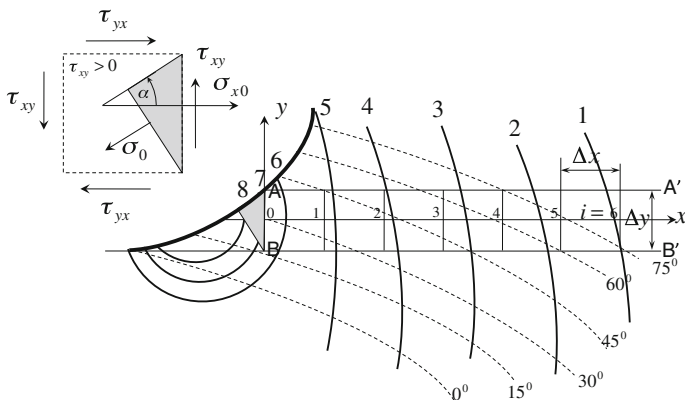


Fig. 3.28 Integration scheme of equilibrium equations at finite differences along the x axis

4. Value and sign of τ_{xy} along the same AA' and BB' .
5. $\Delta\tau_{xy}$ along y .
6. This value multiplied by the ratio $\Delta x/\Delta y$.⁷
7. σ_x along x .
8. $\sigma_x - \sigma_y$ along section $0 - i$ by isochromatics and isoclines values.
9. σ_y from the equation:

$$\sigma_x - \sigma_y = \pm(\sigma_1 - \sigma_2) \cos 2\alpha \tag{3.76}$$

10. Finally, the principal stresses at point i are individually determined:

$$\sigma_{1,2} = \frac{1}{2}(\sigma_x + \sigma_y) \pm \frac{1}{2}(\sigma_1 - \sigma_2) \tag{3.77}$$

3.11.4 Overview of Experimental Methods for Principal Stresses Separation

The experimental methods to separate the principal stresses are based on other theoretical relationships between principal stresses.

3.11.4.1 Determination of the Sum of the Principal Stresses

One is based on the measurement of the third principal strain $\epsilon_3 = \epsilon_z$ perpendicular to the model plane xy , remembering that, in the elastic field for plane stress, the following holds:

⁷ Elements can be selected square, to have a ratio equal to 1 for the all points.

$$\sigma_z = \frac{\nu}{E}(\sigma_1 + \sigma_2)$$

where:

$$\varepsilon_z = \frac{\Delta h}{h}$$

is a linear function of the thickness variation due to the load. It can be measured at any point of the model plane:

- By precision lateral extensometers.
- By optical interferometric methods (such as holographic interferometry, etc.).

3.11.4.2 Method of Oblique Incidence

The fundamental equation of photoelasticity is obtained in the case of light transmission in a direction perpendicular to the model surface. If the plane of the model is rotated so that the direction is no longer perpendicular to the model plane, the isochromatics observed in the polariscope are different from the previous ones and provide additional information. Let us consider for simplicity the case in which the principal directions are known: if the model is rotated around 1 axis of an angle ϑ , Fig. 3.29, the light beam goes through the model with a thickness $h/\cos\vartheta$. Isochromatics are no longer due to stresses lying on the model plane, but to *secondary principal stresses*, i.e. those lying on a plane perpendicular to the propagation direction, such as plane σ_1, σ'_2 :

$$\sigma_1 - \sigma'_2 = f_\sigma \frac{N_\vartheta}{\frac{h}{\cos\vartheta}} \quad (3.78)$$

N_ϑ is the fringe order for the new light beam direction. The link between σ'_2 and σ_2 is:

$$\sigma'_2 = \sigma_2 \cos^2 \vartheta$$

then:

$$\sigma_1 - \sigma_2 \cos^2 \vartheta = f_\sigma \frac{N_\vartheta}{\frac{h}{\cos\vartheta}}$$

Combining Eq. 3.78 with Eq. 3.38 gives:

$$\begin{cases} \sigma_1 = f_\sigma \frac{\cos\vartheta}{\sin^2\vartheta} (N_\vartheta - N_0 \cos\vartheta) \\ \sigma_2 = f_\sigma \frac{1}{\sin^2\vartheta} (N_\vartheta \cos\vartheta - N_0) \end{cases} \quad (3.79)$$

3.12.1 Boundary Effects

When a photoelastic model is stored for a long time even without external loads, a state of stress arises along the external surfaces that causes a spurious birefringence effect, responsible for errors when not identified. This noise is due to water vapor transmission from the air to the material and vice-versa. If the concentration of the water vapor is uniform in the resin, the boundary effect is insignificant; if vice-versa the diffusion process is so slow that it takes a long time to reach an equilibrium state, then the stress developed at the surface and its optical effects are considerable.

The extent of the phenomenon is related to the relative humidity of the environment and to its temperature. The boundary effect also depends on the conditions of the model surface. The machined surfaces show a much more marked effect than surfaces obtained by the casting process.

3.12.1.1 Machining

As regards the machining properties of plastic resins, it should be specified that stresses of thermal origin in the model must be avoided since they tend to be *frozen* together with stresses due to external loads. It is therefore necessary to limit the development of heat and excessive pressure of the tool on the model during the model preparation phase.

Among the machining methods, those in which the cutting edge of the tools does not always remain in contact with the model and allows a partial cooling of the surface, such as band saws, milling machines, and of course lathes, are particularly suitable. In these types of machining it is not necessary to use artificial cooling.

For drilling, the precaution of cooling and using tools with tungsten carbide plate must be observed. For epoxy resins the machining parameters in Table 3.2 are recommended.

Table 3.2 Machining parameters

Tool	Cutting speed (m/min)	Feed speed	Note
Lathe	4–5	0.1 mm/rev	Upper angle clearance: 0–5°
Milling	80–100	2–3 cm/min	End mill
Milling	20	Same	Standard milling
Drill	10–20		With short intervals

3.12.2 Stress Freezing

The behavior of the resin can be explained by a two-phases theory.

According to this theory, the material consists of two structures: a primary elastic and a secondary viscose. The elastic structure is governed by *Hooke's law* and its elasticity modulus is practically independent of temperature; the viscous structure is temperature dependent. The elastic structure made up of a network of long chain molecules of three-dimensional hydrocarbons (*primary bonds*). The viscous structure is made up of other molecules less firmly attached through short chains. When the polymer is at room temperature, both molecular systems, primary and secondary, react to the applied loads.

If the temperature increases up to the critical value (critical temperature), the secondary bond breaks and the primary structure bears all the applied load.

Since the secondary bond is a significant portion of the polymer, the deformations that occur at critical temperature are quite larger than the corresponding ones at room temperature, but in elastic range.

If the temperature is lowered to room temperature with the load applied on the specimen, the secondary bonds are restored around the primary bond that remains deformed and *frozen* in the achieved configuration. If the load is finally removed, the primary system relaxes slightly, but most of the deformation remains inside. Since the phenomenon occurs at the molecular level, the deformation and the birefringence that accompany it are retained in every molecule, in spite of the machining operation, i.e. the model can be cut without modifying the impressed stress and the fringes distribution.

The resolution of the secondary bonds is instantly verified at a temperature greater than or equal to the critical temperature but this phenomenon can happen at lower temperatures, in the case of a longer period of application of the load: the viscosity behavior has time to act and to transfer the entire load to the elastic part. In spite of this fact, at room temperature the same resin has an elastic behavior if the the time of the loads application is short. In this case the elastic modulus is much greater than the elastic modulus of the primary bond only. The frozen stress mechanism is illustrated by the example in Fig. 3.30.

The different properties of the materials used for room temperature tests and for frozen tests are shown in Table 3.3 at room temperature and in Table 3.4 over the critical temperature.

3.12.3 Calibration of Photoelastic Material

The measurement of material parameters necessary to quantify the results of the experiment is obtained by an inverse procedure on a specimen with a known stress state, generally a small bar loaded with bending moment as in Fig. 3.31.

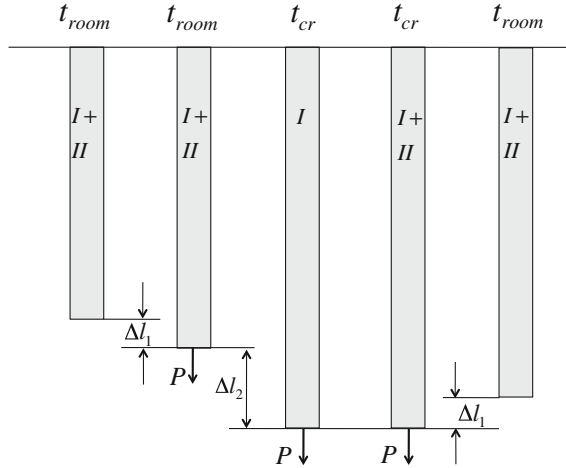


Fig. 3.30 Freezing mechanism for plastic resin, from [9] for a bar loaded by P along its axis: I primary bonds, II secondary bonds, t_{room} = room temperature, t_{cr} = critical temperature, Δl_1 = bar elongation at room temperature, Δl_2 = bar elongation at critical temperature

Table 3.3 Material properties at room temperature

Material	Commer. name units	f_σ ($\lambda = 577$ nm) (N/(mm ord))	E (MPa)	$Q = E/f_\sigma$ (ord/mm)
Resin catalyst	Araldite CT200 + 30 % phthalic anhydride	10.3	3140	305
Resin catalyst ^a	Araldite 6020 + 50 % phthalic anhydride	10.2	3100	305
Polymethyl-methacrylate	Plexiglass Perspex	-130	3000	23
Rubber				
Polyurethane	Hysol 4485	0.17	4	26
Gelatine		0.09	0.3	3.3
Glass		400	70,000	200

^a<http://en.wikipedia.org/wiki/Araldite>

By measurement of the maximum stress at the surface of the bar as a function of the applied bending moment and by the fundamental relationship of photoelasticity the following formula is derived:

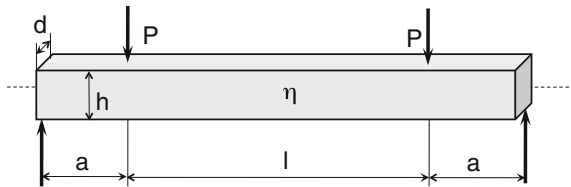
$$f_\sigma \frac{N_{max}}{d} = \frac{M}{W}$$

Table 3.4 Material properties at critical temperature

Commer. name units	f_σ for $\lambda = 577$ nm (N/(mm ord))	E (MPa)	Q (ord/mm)	T (°C)
Araldite CT200 + 30 % phthalic anhydride	0.23	13	56.6	140
Araldite 6,020 + 50 % phthalic anhydride	0.415	36	86.6	162

The ratio $Q = E/f_\sigma$ is called *figure of merit* because it is proportional to the *strain sensitivity*, that clarifies the capacity of the material to be employed for *freezing* better than *stress sensitivity*

Fig. 3.31 Material calibration by a small beam in pure bending



where: $W = \frac{dh^2}{6}$. The expression of photoelastic constant as a function of applied load and maximum order thus becomes:

$$f_\sigma = \frac{6Pa}{h^2 N_{max}} \tag{3.80}$$

The maximum order is usually estimated as the average between the maximum values on the tensile and compression sides of the bar.

The elastic modulus in bending is obtained from the deflection of the central axis, estimated as the average value between intrados and extrados deflections:

$$\eta = \frac{\eta_e + \eta_i}{2}$$

Recalling the expression of the maximum deflection of a beam in bending:

$$\eta = \frac{Ml^2}{8EJ}$$

the elastic modulus is obtained by the following relationship:

$$E = \frac{3Pal^2}{2\eta dh^3}$$

3.12.4 Material Sensitivity

In elastic field the *strain sensitivity* is related to the stress sensitivity by equation:

$$\frac{1}{f_\varepsilon} = \frac{E}{1 + \nu} \frac{1}{f_\sigma}$$

For strict similarity between a prototype and its model the strain should be identical in them:

$$\varepsilon_m = \varepsilon_p$$

Considering a prototype with deformations of the order of magnitude of 1000 $\mu\varepsilon$, looking at the relationship:

$$\varepsilon_1 - \varepsilon_2 = \frac{1 + \nu}{E} (\sigma_1 - \sigma_2) = f_\varepsilon \frac{N}{d}$$

the strain sensitivity $1/f_\varepsilon$ with, e.g. $d = 10$ mm and $N = 2$ (fringe order), should have an order of magnitude of hundreds (200 ord/ $\mu\varepsilon$ mm).

In fact, at *freezing* temperature, the term $1/f_\varepsilon$ never assumes a value greater than 30–40, while it reaches a value of about 200 only in room temperature tests, Tables 3.3 and 3.4.

This observation shows the difficulty encountered in observing a strict similarity between the photoelastic models and metal prototypes.

An approximate similarity can be accepted if the systematic error is estimated. In any case, the ratio $Q = E/f_\sigma$, proportional to the strain sensitivity $1/f_\varepsilon$, is the *figure of merit* of the material.

The allowable stress is assumed to be a very high percentage of the limit of proportionality, which in fact coincides with the ultimate stress. The value of allowable stress allows the choice of the stress scale; once the maximum stress in critical points of the prototype has been estimated, the stress scale is given by $K_\sigma = \sigma_{model}/\sigma_{prot}$ ratio between the allowable stress in the model and the maximum stress in the prototype.

Of course, the maximum stress in the prototype is only an approximate estimation. The actual stress scale will be recalculated exactly at the test end, from the forces and the length really adopted for the model. The preliminary estimation is, however, very useful for choosing the loads that must act on the model in the initial phase of the experiment.

For *Araldite B* e.g., at critical temperature, an allowable stress of about 0.6 MPa should not be exceeded.

3.13 Three-Dimensional Photoelasticity: The Frozen Stress Method

This property of freezing the stress state can be successfully utilized in a special technique developed for analyzing three-dimensional stresses in non-slender bodies, not only on their surfaces but especially at whatever interesting point that can be located inside. Synthetically, the procedure for this technique called *three-dimensional photoelasticity* is the following:

- A scaled model of the prototype is made from photoelastic resin, suitable for freezing stress. (usually epoxy resins).
- Complex models are manufactured by casting. Special care must be devoted to the internal cores that must be removed not at room temperature but at an intermediate temperature between room and critical temperature in order to allow the removal, favored by the larger thermal dilatation coefficient of the resin.
- The model is loaded at critical temperature of the resin in an oven controlled by a thermo-controlled system.
- After a certain time at critical temperature at which only the primary elastic bond is active, the model is cooled very slowly up to room temperature. During this long interval the secondary bonds is reconstituted and stresses become frozen in each molecule of the model, without introducing spurious frozen thermal stresses that have enough time to slow down.
- The external loads are removed at room temperature.
- Once the stress state is frozen, the model is cut into *slices* that can be analyzed in a polariscope as plane models.
- The photoelastic effect is first measured for the stress state in the slice planes.
- Only in special cases (symmetry slices or surface slices cut from plane boundaries) the principal stresses are in these planes.
- For a general slice cut from a three-dimensional body, the principal stresses σ'_1 and σ'_2 are not the real principal stresses but only the stresses lying on this slice plane, related to the first by equilibrium equations. They are called *secondary principal stresses*.

For a beam perpendicular to this plane the photoelastic law gives:

$$\sigma'_1 - \sigma'_2 = f_\sigma \frac{N}{d} \quad (3.81)$$

Similarly, the isoclines on this plane are the locus of points with constant principal secondary stress directions.

The choice of thickness d is subjected to a limitation: it is necessary that the secondary principal stresses on plane xy , maintain, point by point, the same direction and value along the thickness. In practice, between this requirement for which a slice thickness should tend to zero (in order to have a rigorously plane stress state) and the opposite of an optical effect sufficient to be correctly measured, a value of compromise is chosen, ordinarily from 2 to 5 mm.

3.13.1 Limit of Three-Dimensional Photoelasticity

In three-dimensional cases the stress state at a point is given by a stress matrix with six independent terms:

$$\begin{pmatrix} \sigma_x & \tau_{xy} & \tau_{xz} \\ \tau_{yx} & \sigma_y & \tau_{yz} \\ \tau_{xz} & \tau_{yz} & \sigma_z \end{pmatrix} \quad (3.82)$$

Let us fix coordinates axes and the elementary volume dx , dy , dz at a generic point of a three-dimensional model with *frozen* stress state.

Which ones and how many of the *six* terms can be determined by observing in a polariscope the infinitesimal cube in three directions?

For each light beam normal to the three faces of the cube, a first relationship can be derived relating the differences two by two of normal stresses to the orders N and a second relationship between shear stresses and the α angles⁸ which yields:

$$\sqrt{(\sigma_x - \sigma_y)^2 + 4\tau_{xy}^2} = f_\sigma \frac{N_z}{d_z} \quad (3.84)$$

$$\tan 2\alpha_{xy} = \frac{2\tau_{xy}}{(\sigma_x - \sigma_y)} \quad (3.85)$$

$$\sqrt{(\sigma_y - \sigma_z)^2 + 4\tau_{yz}^2} = f_\sigma \frac{N_x}{d_x} \quad (3.86)$$

$$\tan 2\alpha_{yz} = \frac{2\tau_{yz}}{(\sigma_y - \sigma_z)} \quad (3.87)$$

$$\sqrt{(\sigma_x - \sigma_z)^2 + 4\tau_{xz}^2} = f_\sigma \frac{N_y}{d_y} \quad (3.88)$$

$$\tan 2\alpha_{xz} = \frac{2\tau_{xz}}{(\sigma_x - \sigma_z)} \quad (3.89)$$

However, six relationships are not sufficient for a complete determination of the stress state, because only five of them are linearly independent. The six unknowns are in fact:

$$(\sigma_x - \sigma_y), \tau_{xy}$$

⁸ In place of Eqs. 3.85, 3.87 and 3.89 the equivalent relationships expressed by the sines of the angles can be written, e.g. for the plane x, y :

$$\sin 2\alpha = \frac{2\tau_{xy}}{\sqrt{(\sigma_x - \sigma_y)^2 + 4\tau_{xy}^2}} \quad (3.83)$$

$$(\sigma_x - \sigma_z) , \tau_{xz}$$

$$(\sigma_y - \sigma_z) , \tau_{yz}$$

and each of them, (e.g. the difference $\sigma_y - \sigma_z$) is obtained as a linear combination of the other two.

With oblique incidence, measurements of $6N$ values in 6 different directions must be performed in place of $3N$ plus 3α values in the case of normal incidence. It is easy to demonstrate that even the oblique incidence technique obtains 5 independent Equations between the 5 previous unknowns.

The problem presented in this way has only theoretical value, because measurements point by point, even ignoring the impossible cutting operation, lack the advantage of a field analysis, which is a distinctive characteristic of photoelasticity.

3.13.2 Overview of the Mold-Making and Casting Technique

The problems of the preparation of the models are the typical ones of casting artificial resins with high precision shape. Moreover, models must be totally free from states of residual stress; therefore it is necessary to:

1. Assess and exactly maintain the proper casting temperature.
2. Avoid polymerization and thermal origin stresses.
3. Favor the internal cores removal with a suitable mold design and other cares, see Fig. 3.32.

The monomer is heated up to liquid phase; a hardener is added in exact weight proportion. The chemical reaction is exothermic: due to low thermal conductivity of the resin, the temperature may rise locally in the melted condition and cause partial polymerization of some areas, with the consequent formation of spurious internal stress states.

The procedure of pouring molten polymer in preheated mold to 100–125 °C and then of cooling to remove the reaction heat until a uniform temperature has been reached, is recommended. After that, casting is preserved at 80–100 °C.

Figure 3.33 shows the thermal cycle the model must undergo in the mold and the time period for *extracting* the cores which must occur at a temperature higher than the room one to avoid shrinking stresses that would otherwise arise at this temperature for the forcing effect of the resin on the core, due to the particularly high value of the coefficient of thermal expansion of the resin ($\alpha = 6 \times 10^{-5} \text{ C}^{-1}$ versus e.g. $2.5 \times 10^{-5} \text{ C}^{-1}$ for aluminum alloys of casting-mold material). For a finished surface of the model without further processing, it is convenient to use a metal mold (preferably of aluminum alloy) with machined surfaces, coated with suitable anti-adhesives such as silicone oils and greases. Surfaces that do not have special finishing needs or for which the subsequent machining process is foreseen, a simple form of approximate sizes is advantageous.

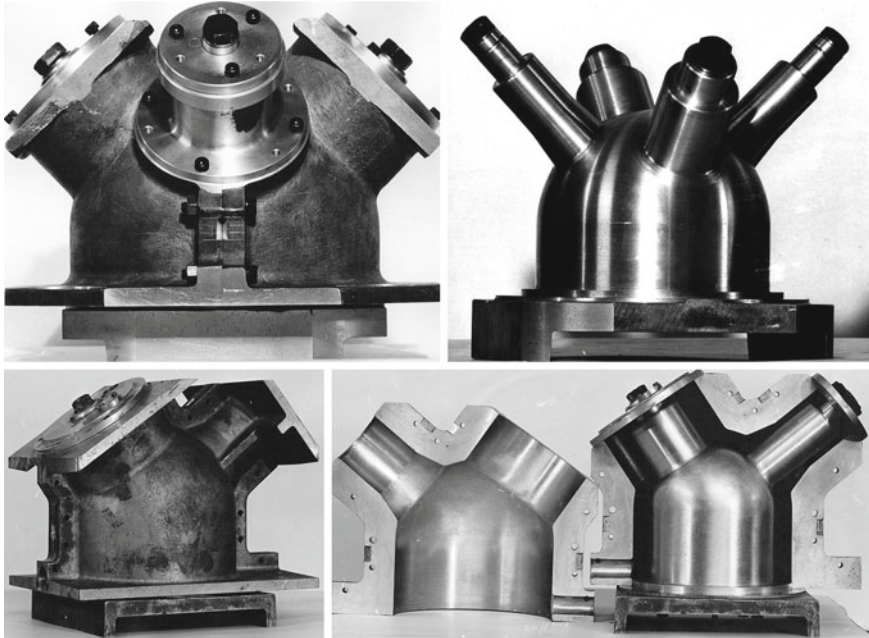
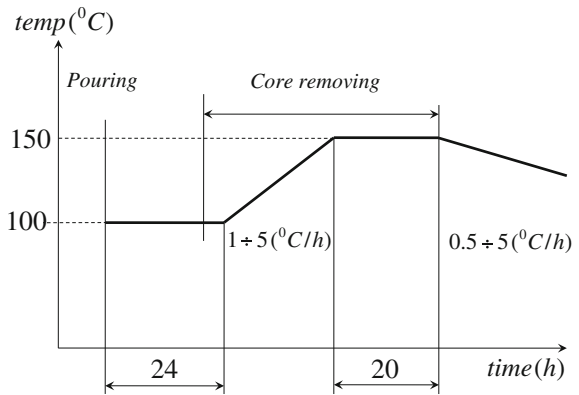


Fig. 3.32 How to build 3D models: the mold design for casting epoxy resin of pressure vessels

Fig. 3.33 Thermal cycle for models casting



If the prototype is available with suitable size, a mold of thermosetting plastic material can be obtained directly from the prototype itself by plaster mold.

3.13.3 Stress State in Plane External Surfaces and in Planes Perpendicular to the External Surface

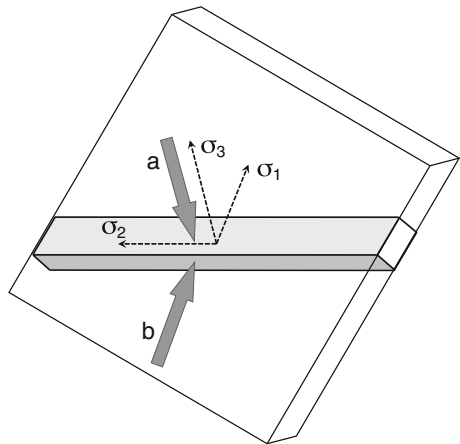
A first criterion for reducing the frozen model to a series of plane models is cutting *slices* containing a portion of plane surfaces and slices perpendicular to external surfaces, (i.e. with edges belonging to the surface). In both cases, the complete determination of stress state is straightforward.

3.13.3.1 Stress State on Plane External Surfaces

In this case in which the model has flat outer surfaces, slices such as those in Fig. 3.34 can be cut. Principal stresses σ_1 and σ_2 are unknown, while σ_3 is zero or known (e.g. applied pressure). In this case:

- A slice is observed with incidence light perpendicular to its plane and the difference ($\sigma_1 - \sigma_2$) and both principal stress directions are derived by isochromatics and isoclines.
- Once the principal directions are known, it is possible to cut *sub-slices* from the main slice with their faces parallel to the plane σ_2 , σ_3 , as in Fig. 3.34 (or σ_1 , σ_3).
- The sub-slice is observed in the direction b of σ_1 (or σ_2) and, by isochromatic values, a further relationship is derived between σ_2 , σ_3 as in the figure (or between σ_1 , σ_3). In this way all the principal stress components σ_1 , σ_2 , σ_3 are determined.

Fig. 3.34 Complete stress determination at every point on a surface slice: from two observations a and b are derived the values of ($\sigma_1 - \sigma_2$) and ($\sigma_2 - \sigma_3$) are derived. If σ_3 is known σ_1 and σ_2 are deduced



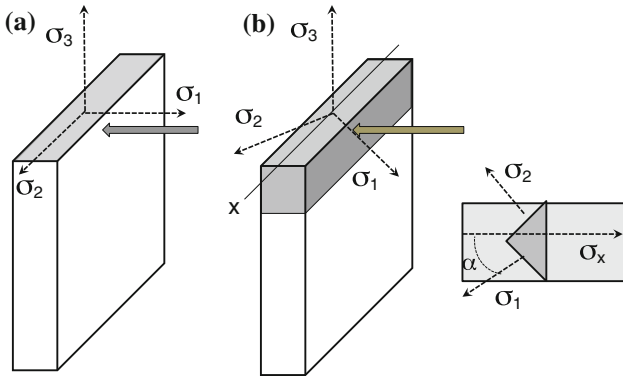


Fig. 3.35 Slice with an edge at the model surface: **a** σ_2 and σ_3 are principal stresses on the mean plane of the slice; **b** σ_3 is the only principal stress on the mean plane of the slice, while σ_1 and σ_2 are out of the plane and σ_x is a secondary principal stress. A sub-slice in gray is observed in the direction of σ_3

3.13.3.2 Slices Containing Edges of External Surfaces

In this case, care must be taken to cut slices with the edges that must be a portion of the model surface with cutting planes perpendicular to the model surface, so that it is possible to conclude the slice contains in its plane one of the three principal directions (e.g. σ_3 as in Fig. 3.35). The determination of stresses at the edge is thus immediate:

1. If the plane is a symmetry plane or otherwise such as to also contain a second principal direction (e.g. σ_2), Fig. 3.35a observing the slice along the other principal direction σ_1 , the difference $\sigma_2 - \sigma_3$ is given directly; if σ_3 is known or equal to zero, σ_2 is determined.
2. If σ_1 and σ_2 have whatever direction Fig. 3.35b and the slice is observed in perpendicular direction, the difference $\sigma_1 - \sigma_x$ is given where:

$$\sigma_x = \sigma_1 \cos^2 \alpha + \sigma_2 \sin^2 \alpha \tag{3.90}$$

The α value, together with a second relationship between σ_1 and σ_2 is obtained by an observation along σ_3 of the *sub-slice* with a face parallel to the plane $\sigma_1 \sigma_2$.

3.13.4 Determination of Stress State at Points Inside the Model

Once the frozen stress state is known at points located on the surface of the model, for example on the edge of a slice perpendicular to the surface, one can extend

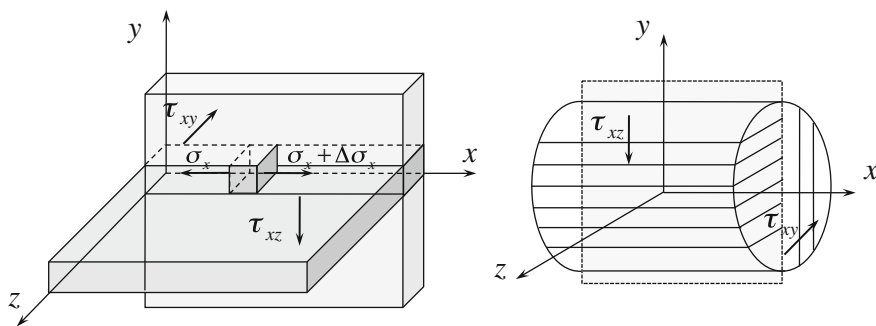


Fig. 3.36 General slices cut from a three-dimensional model on *left side* and when a symmetry plane is present on the *right side*

determination of stress state at points of the slice inside the model. Let x, y, z be a reference coordinate system for the model.

With reference to Fig. 3.36 on the left side, after cutting from the model two slices perpendicular to one another with a common edge along the planes xy and xz , slice xy and slice xz are observed in a polariscope at normal incidence light.

From isochromatics and isoclines, tangential stresses τ_{xy}, τ_{xz} are derived along the common edge of the slices. It is now possible to integrate the equilibrium equation in x direction:

$$\frac{\partial \sigma_x}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} + \frac{\partial \tau_{xz}}{\partial z} = 0 \tag{3.91}$$

$$\sigma_x = \sigma_{x_0} - \int_0^i \frac{\partial \tau_{xy}}{\partial y} dx - \int_0^i \frac{\partial \tau_{xz}}{\partial z} dx \tag{3.92}$$

Similar expressions can be derived for obtaining the other variables σ_y and σ_z , rotating the indexes x, y, z , with the use of other pairs of perpendicular slices.

The following criteria are recommended for slices cutting:

- In the model with a plane of symmetry, slices must be cut in two perpendicular directions shown in Fig. 3.36 on the right side. Shear stresses τ_{xy}, τ_{xz} are measured and the slices are associated with one another and the model can be completely analyzed.
- In the case of asymmetry, the problem of a full determination of the state of stress is much more complex, but the case is more theoretical than practical.
- In order to show the method by means of an application, the example in Fig. 3.37 can be a guide. It refers to a model with frozen stress state due to six vertical concentrated loads on its surface in z direction.

The most interesting unknowns are the stress distribution and their max values $\sigma_z = f(z)$ under each load, along the internal segment \overline{CD} , Fig. 3.37b and the procedure for its determination is the following:

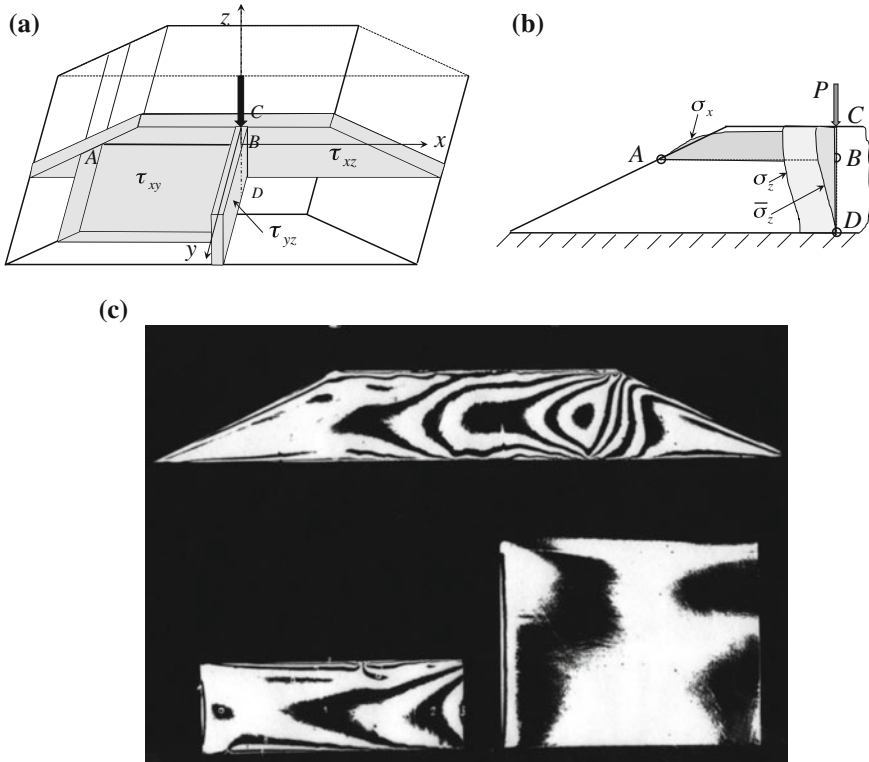


Fig. 3.37 Cutting plan of a three-dimensional model and relative isochromatics images. From an application developed together with Kuske and Robertson [13]

Slices xz, yz Once the coordinate system has been defined, two slices perpendicular to one another with a common edge along z axis are cut from the frozen model, Fig. 3.37a. In Fig. 3.37c the isochromatics patterns for both slices are shown.

$\bar{\sigma}_z = f(z)$ Along the common edge the integration for σ_z (previously described for for σ_x), now specialized in Eqs. 3.93 and 3.94, must be developed e.g. from point D to point C in z direction. From the general expression of the third equilibrium equation:

$$\frac{\partial \sigma_z}{\partial z} + \frac{\partial \tau_{zx}}{\partial x} + \frac{\partial \tau_{zy}}{\partial y} = 0 \tag{3.93}$$

the following integration for finite differences is developed:

$$\sigma_z = \sigma_{z0} - \int_0^i \frac{\partial \tau_{zx}}{\partial x} dz - \int_0^i \frac{\partial \tau_{zy}}{\partial y} dz \tag{3.94}$$

The values of τ_{xz} and τ_{yz} , necessary for the the integration along z , are given by the observation in the polariscope of both slices that, by isochromatics and isoclines give the tangential stresses.

σ_{zD} unknown The difficulty that immediately appears is due to the unknown values of the boundary condition on σ_z at point C or point D , both subjected to external loads. For this reason the diagram of σ_z along z is determined except for the unknown value of σ_{zD} . This incomplete diagram is indicated in Fig. 3.37b with the symbol $\overline{\sigma_z} = f(z)$

Slices xz, xy In order to overcome this difficulty, a third slice #3 perpendicular to the previous ones is extracted from the model parallel to x, y plane. The new common edge between the slice #1 and the slice #3 allows a second integration from point A to point B . At point A lying on the external free surface, the stress σ_{x0} is known. The values of τ_{xz} and τ_{xy} necessary for the integration are obtained by the observation of slices #1 and #3 that, through their isochromatics and isoclines, give the tangential stresses and allow the complete integration for obtaining $\sigma_x = f(x)$. Thus, the σ_{xi} in point $i = B$ is determined.

$\sigma_z = f(z)$ Once σ_{xB} has been determined, $\sigma_{xB} - \sigma_{zB}$ is derived from the isochromatic order and isocline at point B . The real σ_{zB} value can now be immediately obtained. With this value all the diagram $\overline{\sigma_z} = f(z)$ can be translated until this value at point B and, consequently, the real diagram is obtained together with σ_{zC} and σ_{zD} unknown values

As it was observed, photoelasticity continues to be a useful method for stress analysis, in spite of its limitations. Just to quote some classic and some recent studies that utilize this technique, the reader can see the following papers on this topic: [19–22] regarding the influence of thermal cycling on residual stress generated in glass using digital photoelasticity [23], regarding force analysis in porous media [24], regarding temporal phase unwrapping works at stress concentration zones [25], regarding novel photoelastic materials and methods for validation of biomechanics models [26], regarding stress singularity and neutrality around a stiffener revealed by photoelasticity [27], regarding analysis of obliquely oriented edge cracked semicircular rings (OECSR) [28], regarding photoelastic study of an inclined crack from a fillet of a biaxially loaded cruciform specimen. In [29] an application is presented about processing the behavior of particulate materials using photo stress analysis technology that will be introduced in the following section. All these papers were presented at the ICEM Conference in Cambridge in Summer 2014.

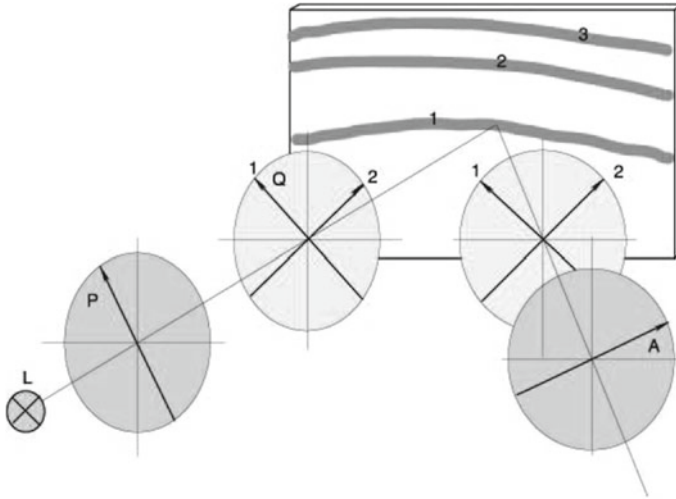


Fig. 3.38 Set up of optical filters in a reflection polariscope. Unlike the transmission polariscope, the axes directions of the quarter-wave filters with the same names are parallel, being crossed by light beams in opposite directions, since the light passes back and forth

3.14 Birefringent Coating Method

The photoelastic coatings make it possible to determine the difference between the principal strains on the surface of a loaded non transparent body. This is an interesting extension of the photoelastic capability in order to investigate directly on prototypes or on not transparent bodies; it is sufficient that a birefringent coating can be cemented on the body surface with an adhesive that has reflective properties. Similarly to electrical strain gages, surface strains of the body (*not stresses!*) are transmitted identical to the photoelastic coating and induce a temporary birefringence in the coating which is measured in a *reflection polariscope*, a simple variant of the classic one.

The photoelastic technique is still used to measure the interference of polarized light, as long as the value of thickness d in the transmission polariscope formula is replaced by $2d$ (since the light passes now back and forth through the coating thickness), and the polariscope is used with a set-up shown in Fig. 3.38, in order to allow the reflection of light on the reflective layer of the cement (Fig. 3.39).

3.14.1 Measurement of Principal Strains Directions

The principal strains directions refer to a Cartesian coordinate system, consisting of the symmetry axes of the body or, in their absence, of horizontal and vertical directions.

Fig. 3.39 Self-manufactured reflexion polariscope



In the same way as the observation in a *transmission polariscope*, the isoclines are observed in white light because here they appear black on a colored background of isochromatics.

At every point of an isocline, the principal strains directions are parallel to the crossed axes of the polarizer and analyzer P and A . The isocline value at a point of the coating is obtained by the joint rotation of both polarizer and analyzer, until the isocline is brought to that point. The value of the isocline is given by the angle of rotation of the filters with respect to the predetermined reference position. From the isoclines obtained for a number of selected directions, the isostatics family can be deduced as previously described.

If the isoclines appear sharp, the directions of the principal strains vary greatly from one point to another; if the isoclines are large and shaded, the directions vary little, as happens, e.g. in a rectangular specimen loaded along its axis, with small geometrical discontinuities.

3.14.2 Strains and Stresses Measurement

When a body with the coating on the surface is loaded, the surface deformations are transmitted to the coating. The relationship between the principal stresses difference

and the fringe order in the coating is given by Eq. 3.38, with the following adjustment:

$$\sigma_1 - \sigma_2 = \frac{N}{2d} \cdot \frac{\lambda}{B} \quad (3.95)$$

Remembering the expression for the coating:

$$\varepsilon_1 - \varepsilon_2 = \frac{1 + \nu'}{E'} (\sigma'_1 - \sigma'_2) \quad (3.96)$$

(apices refer to the coating variables and the absence of apices to the underlying body), the following relationship holds:

$$\varepsilon_1 - \varepsilon_2 = \frac{1 + \nu'}{E'} \cdot \frac{N}{2d} \cdot \frac{\lambda}{B} = \frac{N}{2d} \cdot \frac{\lambda}{K} = N \cdot f \quad (3.97)$$

where:

$$\frac{1}{K} = \frac{1 + \nu'}{E'} \cdot \frac{1}{B} \quad (3.98)$$

and

$$f = \frac{\lambda}{2d} \cdot \frac{1}{K} \quad (3.99)$$

After obtaining strains in the coating, the stresses on the body surface are deduced by Eq. 3.96; in the formula, however, the difference of the principal stresses, the elastic modulus and the Poisson's ratio of the material are those of the body (and not of the photoelastic layer):

$$\sigma_1 - \sigma_2 = (\varepsilon_1 - \varepsilon_2) \cdot \frac{E}{1 + \nu} = N \cdot f \cdot \frac{E}{1 + \nu} \quad (3.100)$$

In many applications one of the principal stresses is zero. In these cases the simplified expression valid for the other non-zero principal stress is:

$$\sigma = N \frac{f \cdot E}{1 + \nu} \quad (3.101)$$

3.14.3 Coating Calibration

The isochromatic fringes observed in the circular polariscope in white light appear colored. The zero-order fringe looks black and makes it possible to assign the proper orders to other isochromatics.

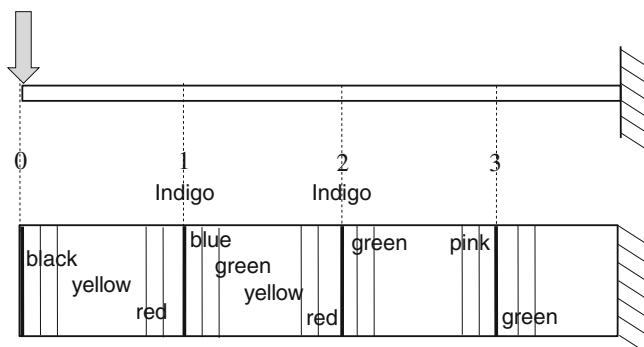


Fig. 3.40 Fringes sequence on calibration bar

A rectangular cantilever beam of the same material as the body, loaded by a known load, with the same type and same thickness of coating used on the body, Fig. 3.40 is used for coating calibration according to Eq. 3.99.

If the color of the transmitted light is *yellow* with a wavelength equal to:

$$\lambda = 576.6 \text{ nm} \quad (3.102)$$

the constant f becomes:

$$\frac{\lambda}{2dK} = \frac{288.3}{dK} \quad (3.103)$$

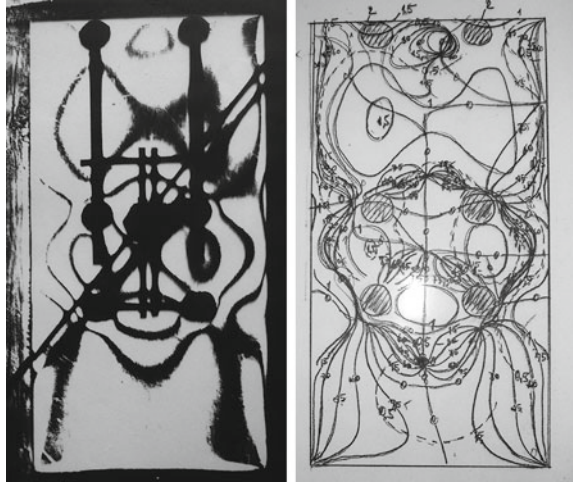
Order $N = 1$ is associated with the fringe with *indigo-violet* color, a complementary of yellow.

Order 1 is thus located where in white light the indigo-violet color is located, i.e., at the passage between warm and cool colors (the end of red, the beginning of blue) and similarly the highest orders from $N = 3$ are located between red and green, because for them the blue color disappears.

For a correct assignment of the fringe orders with increasing strains, colors are always in the sequence *yellow-red-green*.

The measurement of fractional order must be performed by the *Tardy method*. Figure 10.9 in Chap. 10 shows an example of application of reflexion photoelasticity to a double-curvature body. In this case a special technique is applied to adapt the coating in a semi-cured state to the body, before cementing it [30].

Fig. 3.41 Example of application of reflexion photoelasticity to a plane sheet simulating the surface layer in the model of Fig. 3.37, loaded at six points. From an application developed together with Kuske and Robertson [13]



The method provides field information on the surface strain of loaded bodies (Fig. 3.41), whether they are models or real components, plane or with double curvature.

With this method it is possible to identify:

- High stressed zones.
- Zones with a high stress gradient (with dense fringes).
- Scarcely loaded zones.
- Zones with low stress gradients (uniform colors).

The results are of great value for shape optimizing of structural components, especially for automotive and other light structures.

The products needed for the coating, adhesives and the procedure for preparing curves surfaces that match complex shapes, are provided by bulletins of specialized companies, like the one cited previously. The choice must be made on the basis of expected strain and of the desired thickness value which must be such that it does not create appreciable reinforcement effects and sufficient optical sensitivity i.e. the number of fringes wanted for a given level of stress state ($K = 0.001-0.009$).

References

1. Ajovalasit A, Barone S, Petrucci G (1998) A review of automated methods for the collection and analysis of photoelastic data. *J Strain Anal* 33(2):75–91
2. Asundi A (1998) Recent advances in photoelastic applications. School of Mechanical and Aerospace Engineering. http://www.ntu.edu.sg/home/masundi/optical-methods/photoelasticity/recent_advances.html
3. Diaz FA, Patterson EA, Siegmann P (2009) A new experimental method for calculating KI and KII using photoelasticity. In: SEM (ed) Proceedings of the 2009 annual conference, SEM, Bethel
4. Patterson EA (2002) Digital photoelasticity: principles, practice and potential. *Strain* 38(1):27–39
5. Ramesh K (2000) Digital photoelasticity: advanced techniques and applications, vol 1. Springer, Berlin
6. Ajovalasit A (2009) Analisi sperimentale delle tensioni con la fotomeccanica: fotoelasticita, moire, olografia, speckle, correlazione immagini, vol 1. Aracne Roma, Roma
7. Polaroid filter on Wikipedia (2014). [http://en.wikipedia.org/wiki/Polaroid_\(polarizer\)](http://en.wikipedia.org/wiki/Polaroid_(polarizer))
8. Jones RC (1941) A new calculus for the treatment of optical systems. *J Opt Soc Am* 31(7):488–503
9. Dally JW, Riley WF (2005) Experimental stress analysis, 4th edn. McGraw-Hill, New York
10. Föppl L, Mönch E (1959) Praktische Spannungsoptik, vol 1, 2nd edn. Springer, Berlin
11. Frocht MM (1948) Photoelasticity, vols 1 and 2. Wiley, New York
12. Jessop HT, Harris FC (1949) Photoelasticity, vol 1. Hume Press, London
13. Kuske A, Robertson G (1974) Photoelastic stress analysis, vol 1, 1st edn. Wiley, London
14. Mondina A (1958) La Fotoelasticità, vol 1. Rivista di Meccanica, Milano
15. Wolf H (1961) Spannungsoptik. Springer, Berlin
16. Nelson SA (2014) Earth and environmental sciences 2110 mineralogy. Syllabus 1, Tulane University, Tulane, USA, <http://www.tulane.edu/sanelson/eens211/index.html>
17. Post D (1966) Fringe multiplication in three-dimensional photoelasticity. *J Strain Anal Eng Des* 1:380–388
18. Kenny B, Huggill JW (1967) Comparison of fringe multiplication and Tardy compensation methods for a frozen stress investigation. *Strain* 3(1):3–7. doi:10.1111/j.1475-1305.1967.tb00857.x, <http://dx.doi.org/10.1111/j.1475-1305.1967.tb00857.x>
19. Borbas L (2004) Some problems of data evaluation of photoelastic coating technique in case of small-size, fibre-optics fitted equipment. In: Advances in experimental mechanics, vol 1. ICEM 12, Bari, McGraw-Hill, pp 595–596. ISBN 88 386 6273-8
20. Borbas L, Thamm F, Olah L (2006) Comparison of strain gage technique and photoelastic coating method in the investigation procedure of femur prostheses. *J Comput Appl Mech* 7(1):3–12
21. Gyökös F, Borbas L, Thamm F (2005) Following the process of failure during fatigue testing of a bus frame structure by photoelastic coating, vol XII, Slovakian Academy, Zilina, pp 7–12. ISSN 1335-0803
22. Ramakrishnan V, Ramesh K (2014) Influence of thermal cycling on residual stress generated in glass using digital photoelasticity. In: Proceedings of ICEM 16, EURASEM and BSSM, Cambridge UK, see <http://www.icem16.org/contenu.php?page=prog>
23. Gotkhindi T, Simha KRY (2014) Force analysis in porous media. In: Proceedings of ICEM 16, EURASEM and BSSM, Cambridge UK, <http://www.icem16.org/contenu.php?page=prog>
24. Huang MJ, Huang ZW (2014) Temporal phase unwrapping works at stress concentration zones. In: Proceedings of ICEM 16, EURASEM and BSSM, Cambridge UK, <http://www.icem16.org/contenu.php?page=prog>
25. Tomlinson R, Taylor Z (2014) Novel photoelastic materials and methods for validation of biomechanics models. In: Proceedings of ICEM 16, EURASEM and BSSM, Cambridge UK, <http://www.icem16.org/contenu.php?page=prog>

26. Noselli G, Bigoni D et al (2014) Stress singularity and neutrality around a stiffener revealed by photoelasticity. In: Proceedings of ICEM 16, EURASEM and BSSM, Cambridge UK, <http://www.icem16.org/contenu.php?page=prog>
27. Surendra KVN, Simha KRY (2014) Analysis of obliquely oriented edge cracked semicircular ring (OECSR). In: Proceedings of ICEM 16, EURASEM and BSSM, Cambridge UK, <http://www.icem16.org/contenu.php?page=prog>
28. Shreesudha B, Abhinav T, Ramesh K (2014) Photoelastic study of an inclined crack from a fillet of a biaxially loaded cruciform specimen. In: Proceedings of ICEM 16, EURASEM and BSSM, Cambridge UK, <http://www.icem16.org/contenu.php?page=prog>
29. Antony SJ, Barakat T (2014) Processing behaviour of particulate materials using photo stress analysis technology. In: Proceedings of ICEM 16, EURASEM and BSSM, Cambridge, UK, <http://www.icem16.org/contenu.php?page=prog>
30. Photoelastic coating materials (2014). <http://www.vishaypg.com/micro-measurements/photo-stress-plus/category/coating/?subCategory=materials>

Chapter 4

Introduction to Holographic Interferometry

Abstract Holographic interferometry is a method for recording and reconstructing three-dimensional images, which can be applied for the solution of strain analysis problems. Today this method, at least in its classical form, suffers from the way in which the interference fringes are recorded. Special photographic plates with high resolution are needed to record these images, together with accurate manipulation. This constraint has limited the industrial application of the method and oriented the research towards other optical interferometric methods that are less demanding in terms of equipments and experimental difficulties. Nevertheless, classic holographic interferometry offers an invaluable tool for the high accuracy and resolution required by very small displacements analysis. In the present book the classic method is presented in order to understand a special application to Fracture Mechanics for parameters identification in the case of interaction between two cracks.

4.1 Holography

A coherent light beam that hits the surface of an object and is diffused by it until it reaches a recording system, as special photographic plates, retains information of the distance from the plate of every points on the surface of the object, i.e., of its three dimensional shape. Holography is based on this principle. In the case of an object deformation, an overlap is done of undeformed and deformed images of the object. Two coherent light paths coming from undeformed and from deformed images of the object interfere with each other. For the smallness of the displacements in comparison to the size, the two images are indistinguishable, but the two light waves, no longer coincident, generate a system of interference fringes that appear on the object's surface. They store information on distances between the two configurations, i.e. on the displacements of every point of the surface.

The light radiation presented in the Chap. 3 is called *coherent*, when a defined relationship between phases of the light vectors exists during the time and along the direction of propagation, (*temporal* coherence and *spatial* coherence). Natural light is neither temporally coherent, because the emission occurs by irregular pulses and

not in continuous wave trains at constant frequency, nor spatially coherent, since each point of a source emits radiations without phase relationship with the radiations of other points.

4.1.1 The Laser Light

Today, *Laser*¹ is utilized as a light source with high spatial and time coherence, that has allowed the development of a variety of optical techniques with many applications in different measurements fields. Laser, in addition to high coherence, has a high luminous flux, which enables images recording at high resolution, essential for *holograms* generation [4–7, 9, 11, 12]. The coherence length of laser used in holography applications, i.e. the maximum distance along the direction of propagation for which a defined phase of the light radiation is maintained, is of about half a meter. For the applications of coherent optics in structural problems, the following laser types are utilized:

- *He-Ne, helium-neon* that emits at a wavelength of 632.8 nm (red light), with a radiation power from 5 to 50 mW.
- *Argon* laser, that emits at wave length of about 514 nm (green light), with a power from 100 mW to 5 W.
- *Ruby* laser is utilized in dynamic applications, for its capacity to emitting pulsed light with pulses of short duration (10^{-8} s), with power from 100 kW to 10 GW ($1 G = 10^9$), and a wave length of 690 nm.

Laser emits light beams with diameter of about one millimeter from which plane or spherical wavefronts are achieved with appropriate optical devices.

The laser light is also plane polarized, a prerequisite (together with monochromatism) of spatial coherence; so it is possible to apply the concepts of constructive and destructive interferences, discussed in Chap. 3. Furthermore, in order to maintain a defined phase relationship between two waves, it is necessary for them to be radiated by the same source, since two individually coherent sources are not mutually coherent.

4.1.2 Hologram

Let us consider a laser and the optical system shown in Fig. 4.1. A lens, or rather an objective (generally 10x or 20x), transforms the Laser parallel beam to a spherical wave. This optical element is called a *spatial filter* because it combines a calibrated hole of small size positioned in its focus with the objective, with the task of filtering spurious light components. A mechanical system mounted on slides with micrometer screws, gives exact positioning of the hole into the focus.

¹ Acronym for *Light Amplification by Stimulated Emission of Radiation*.

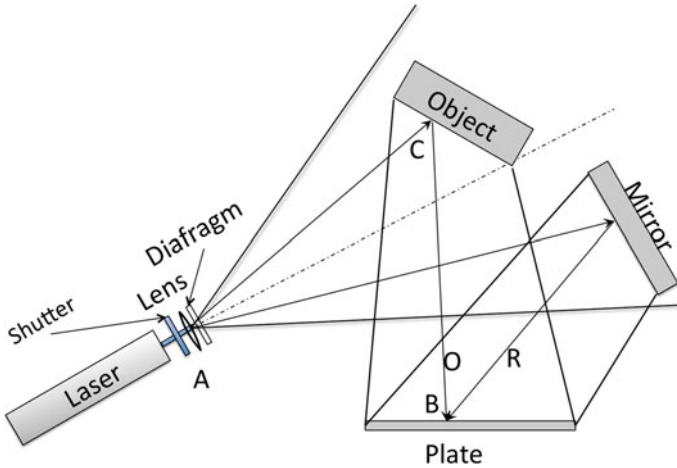


Fig. 4.1 Layout of the optical system for the creation of hologram

With the most simple set-up, a portion of the spherical wave is reflected directly from a mirror on the photographic plate and constitutes a reference field. A second part of the spherical wave illuminates the object, each point of which scatters light in every direction. The portion of the diffused light, collected by the photographic plate, is called the *object field*.

Thus the simultaneous presence of the reference field and of the object field are present on the plate where they optically interfere, generating high density interference fringes.²

After photographic development, the plate constitutes an hologram that contains information on amplitude and phase of the light coming from the object. The fineness of the grain of a photographic plate is necessary for holograms generation and is not a common commercial characteristics. It must have the resolution of 3,000–5,000 lines/mm.³ The complex amplitudes of the light coming from the object and from the mirror at a point of the plate (point B), are O and R [8]:

$$O = |O| e^{i\omega} \qquad R = |R| e^{i\rho} \qquad (4.1)$$

then the light intensity at point B is, see Sect. 3.4 in Chap.3:

$$\begin{aligned} I &= (O + R)^*(O + R) = |O|^2 + |R|^2 + |O| \cdot |R| e^{i(\omega-\rho)} + |O| \cdot |R| e^{-i(\omega-\rho)} \\ &= |O|^2 + |R|^2 + 2|O| \cdot |R| \cos(\omega - \rho) \end{aligned}$$

² For a correct exposure of the recording plate in darkened room, the average intensity of both beams must be similar.

³ Today these special silver halide holographic plates are not easy to find for the end-production by some Companies. The holographic techniques aimed at creating media for holographic data recording are sometime announced by companies in the electronics industry such as Optware, Fujifilm, CMC Magnetics.

The intensity is a real quantity, a function of phase angle ω of the light coming from the object.

This is due to the presence of the reference beam, because, without it, the intensity would be only related to the square of $|O|$ and would not be dependent on phase ω .

The phase is related to the distance between B and all points of the object visible from the point B : each point in the hologram contains information about its distances from all the points of the object, in other words, on the three-dimensional form of the object.

It is understandable that the coherence of the light used is necessary in order to guarantee the constancy of $|O|$, $|R|$, ω and ρ and consequently the positioning of the optical system on a non-vibrating support is an imperative specification. To obtaining a hologram it is necessary that all elements of the system do not experienced relative displacements higher than fractions of the wavelength of the used light ($1/8-1/10 \cdot \lambda$) in order that the phase angles ω and ρ remain unchanged.

Let us consider the hologram produced by a photographic plate. The exposure, i.e. the product of the intensity of incident light by the time of exposure, produces a developed photographic plate that is made up of very fine black and bright fringes. If the plate is then crossed by a light beam, it acts as a filter with transmission function T variable from point to point. This transmission function, for each ray that crosses the filter, is defined as the ratio between the transmitted and incident complex light amplitudes:

$$T = \frac{A_t}{A_i} = \frac{|A_t|e^{i\phi}}{|A_i|e^{i\phi}} = \frac{|A_t|}{|A_i|} \quad (4.2)$$

This filter modifies only amplitude but not the light phase.

The transmission function T of the hologram is proportional, at each point, to the exposure that is the product $I \cdot t$, where I is the intensity of light that impressed the plate at that point and t is the exposure time:

$$T = T_0 - b \cdot (I t) = T_0 - kT \quad (4.3)$$

where T_0 and b are constants dependent on the emulsion quality and development conditions, Fig. 4.2.

4.1.3 Reconstruction

If the photographic plate (developed and fixed) is repositioned in the plate-holder in the exact position of the exposure and illuminated with only the reference beam (removing the beam coming from the object), the following complex amplitude A_t is recorded at point B :

$$\begin{aligned} A_t &= R \cdot T = R (T_0 - kT) \\ &= |R| \cdot [T_0 - K (|O|^2 + |R|^2)] \cdot e^{i\rho} - K|O||R|^2 e^{i\omega} - K|O||R|^2 e^{i(2\rho-\omega)} \end{aligned} \quad (4.4)$$

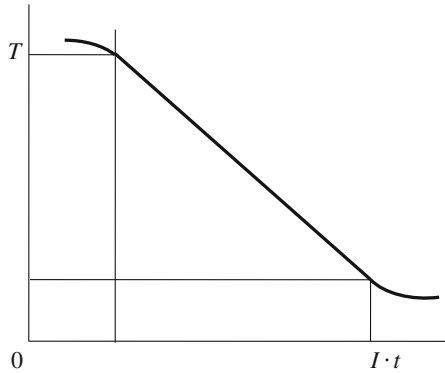


Fig. 4.2 Transmission function of a photographic plate versus exposure energy

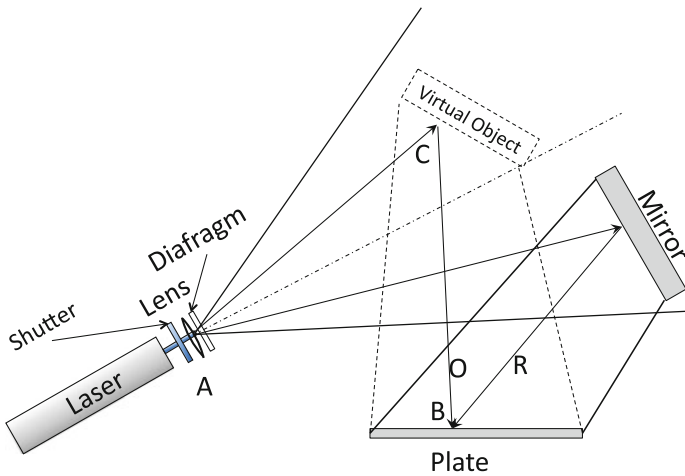


Fig. 4.3 Reconstruction for hologram generation

The term $-K|O||R|^2e^{i\omega}$ has the same phase and modulus proportional to the complex amplitude modulus of the object beam $|O|$. The negative sign shows that the vector has a opposite direction of the object vector.

From every point of the hologram it is then possible to see the three-dimensional object image in the same position that the real object had, Fig. 4.3; other terms represent only noise effects but do not significantly deteriorate the quality of the reconstructed object. The quality of the object image depends on the extent of the surface of the hologram, even if only one point of the hologram receives and then records in modulus and phase the rays diffused from all the points of the object. Each small fraction of the hologram provides the information for the complete object reconstruction.

4.2 Holographic Interferometry

If the reconstruction of the hologram is performed by repositioning the developed plate and also the object in the recorded locations, an observer positioned behind the plate receives two waves, the object and the reconstructed waves, perfectly coincident and indistinguishable.

If the object is loaded, the displacements produced in the body slightly change the object's shape. For the smallness of the displacement with respect to the size, the object and its holographic image are indistinguishable, but the relative light waves, no longer coincident, interfere, generating a system of fringes that appear on the surface object.⁴ There are alternative arrangements of the optical elements more complex than the one outlined here but more efficient from the point of view of the exploitation of the laser power.

One of these is shown in in Fig. 4.4. The light beam is split at the Laser output by a semi-reflective mirror; the first beam, through a spatial filter, is projected on the object, while the other, through a second lens, is conveyed directly on the recording plate as reference beam.

Holographic interferometry can be classified into the following three categories:

1. Double-exposure holographic interferometry.

With a double exposure, both holographic images of undeformed and deformed body are recorded on the photographic plate. They interfere, generating fringes that appear *frozen* on the body surface in the reconstruction phase when the real body is taken away from the optical path.

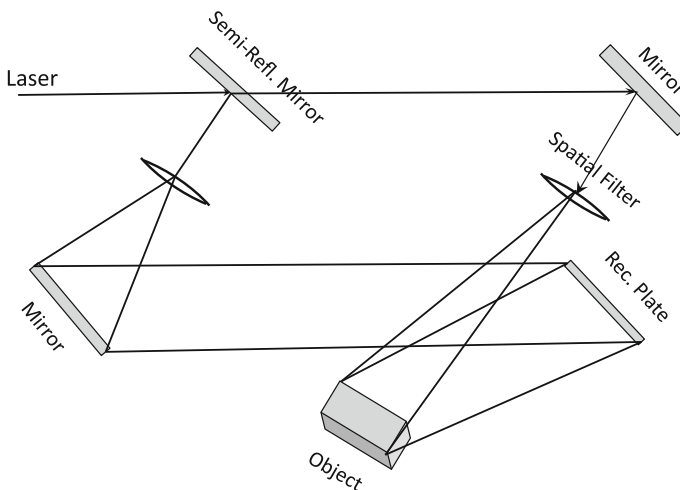


Fig. 4.4 Most common arrangement of the optical system for holograms generation

⁴ Of course any rigid movement and rotation is equally recorded and overlapped to the displacements due to the loads.

With this procedure the plate is exposed twice, overlapping two different holograms: a first with an undeformed object and a second with a deformed object. Only two body configurations are compared and analyzed but the method offers the considerable advantage of not requiring the accurate repositioning of the developed photographic plate in a rigid plate holder.

2. *Holographic interferometry in real time.* In the observation phase only the hologram of non-deformed body is impressed on the recording plate. The interference fringes in the reconstruction phase are visible if this hologram and the real image of the deformed object are accurately overlapped.

Fringes vary with the instantaneous deformation (and rigid movements if present) of the body, as a function of external loads. The procedure requires accurate repositioning of the developed photographic plate on the stable plate holder where the first hologram was generated.

The advantage is the possibility to observe different levels of body deformation with only one hologram of the undeformed state.

3. *Holographic interferometry in time average.* This is used for the amplitudes recording of vibrations of a vibrating body; with this procedure the hologram is recorded for an exposure time greater than the vibration period, so that only the two extreme configurations of the body at zero vibration speed interfere. Only fringes at maximum vibration amplitude are recorded, while the other positions with no-zero vibration velocity are not recorded on the photographic plate. This method is specific for the vibration analysis of thin beams and plates. In a first approximation, since the vibrations have cancelled the intermediate positions, the hologram is equivalent to the double exposure of the extreme displacements of the body at zero speed.

4.2.1 Interference Fringes in a Double Exposure Hologram

Let us consider the displacement vector from a point C of the object—assuming that the displacement is contained on the plane defined by three points A , focus of the spatial filter, C generic point on the body and B one observation point, fixed on the photographic plate, Fig. 4.5—and the coherent light path coming from the point C before and after the deformation.

Since both light vectors come from two reconstructed images, they are of course plane polarized but also vibrating in the same direction. Thus, the constructive interference condition (*bright fringes*) between the beams before and after the deformation is obtained by equating the path difference with an integer number of wavelengths,

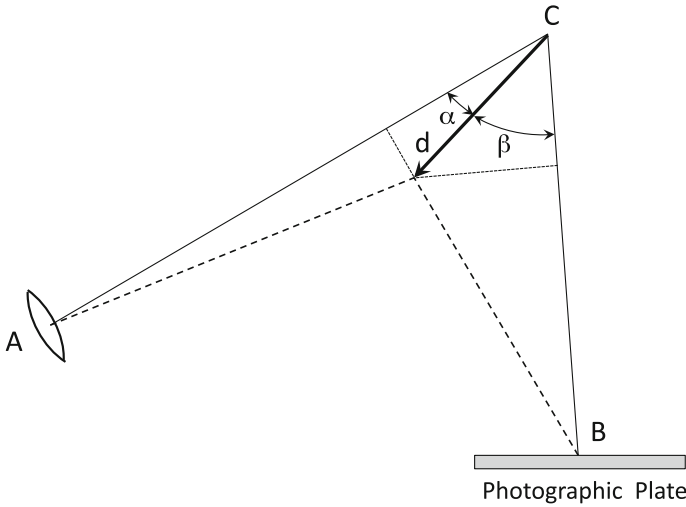


Fig. 4.5 The two paths of light beams at point *C* before and after the displacement *d* (with an amplified scale for displacements). See also Eq.4.11

while the destructive interference condition (*black fringes*) is obtained by equating the path difference with an odd number of semi-wavelengths, Eq.4.5:

$$d (\cos \alpha + \cos \beta) = N\lambda = \begin{cases} n \lambda & \text{bright fringes} \\ \frac{2n + 1}{2} \lambda & \text{black fringes} \end{cases} \quad (4.5)$$

Displacement *d* can be deduced by the fringes orders when, in addition to the lying plane, its direction is also known:

$$d = \frac{N \lambda}{\cos \alpha + \cos \beta} \quad (4.6)$$

with:

- $N = 0, 1, 2, 3, \dots$ for bright fringes
- $N = 0.5, 1.5, 2.5, \dots$ for black fringes.

4.2.1.1 Null Intensity Condition

An alternative way to obtain Eq.4.5 is given respectively by the condition of maximum light intensity or of light extinction (null intensity) at point *C*.

Denoting by $A_1 = |O| e^{i\rho_1}$ and $A_2 = |O| e^{i\rho_2}$ the two waves coming from the undeformed and deformed reconstructed images and Δ the phase difference between the two waves $\rho_2 - \rho_1$, the total intensity is:

$$\begin{aligned} I &= |O|^2 (e^{i\rho_1} + e^{i\rho_2})^* \cdot (e^{i\rho_1} + e^{i\rho_2}) = |O|^2 (e^{-i\rho_1} + e^{-i\rho_2}) \cdot (e^{i\rho_1} + e^{i\rho_2}) \\ &= |O|^2 \cdot [2 + e^{i\Delta} + e^{-i\Delta}] \end{aligned}$$

then:

$$I = 2|O|^2 \cdot (1 + \cos \Delta) \quad (4.7)$$

The extinction condition is obtained when:

$$\cos \Delta = -1$$

i.e.:

$$\Delta = \pi + 2n\pi$$

N is thus an integer number for the bright fringes, while for the black fringes:

$$N = \frac{\Delta}{2\pi} = \frac{1 + 2n}{2} = 0.5 \quad 1.5 \quad 2.5 \quad \dots \quad (4.8)$$

which are identical to Eq. 4.5.

4.2.2 Interference Fringes in a Real Time Hologram

This case differs from the previous one, since the interference fringes arise from the comparison of a light vector coming from the reconstructed image of the undeformed object and the one from the real deformed object.

It was observed in Eq. 4.4 in 4.1.3 that the term with the same phase and modulus proportional to the complex amplitude modulus of the object beam $|O|$ has opposite direction to the object vector.

Consequently, in this case it is the destructive interference condition (*black fringes*) between the beams before and after the deformation that is fulfilled when the path difference is equal to an integer number of wavelengths.

The condition of Eq. 4.7 changes in this way:

$$I = 2|O|^2 \cdot (1 - \cos \Delta) \quad (4.9)$$

and the extinction condition becomes: $\cos \Delta = 1$ i.e.:

$$\Delta = 0 + 2n\pi$$

with n integer. The sequence of orders N for black fringes becomes:

$$N = \frac{\Delta}{2\pi} = n = 0 \quad 1 \quad 2 \quad 3 \quad \dots \tag{4.10}$$

It is easy to observe the analogy between constructive and destructive interferences between the light vectors in the holographic interferometry and between vectors coming from the analyzer filter of a polariscope in photoelasticity.

The double exposure hologram gives the same interference condition as parallel polarizer and analyzer filters with white fringes of integer orders, while the real time hologram is the same as the crossed filters, with black fringes of integer order.

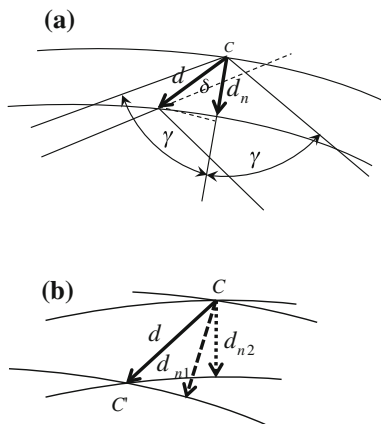
4.3 Description by Means of Ellipses

It may be remarked [4], that the same optical path difference and therefore the same fringe order are common to all possible displacements of the point C represented by vectors outgoing from C with the other extremes on the *ellipse* with foci A and B , such that the distance d_n between the point C and the ellipse, Fig. 4.6, is equal to:

$$d_n = \frac{N\lambda}{2 \cos \gamma} \tag{4.11}$$

where 2γ is the angle between the illumination and observation beams. The difference between Eqs. 4.6 and 4.11 is clear: in Eq. 4.11 the angle 2γ is known and d_n can be

Fig. 4.6 **a** Intersections between the ellipses and normal component d_n of the real displacement d of the point C . **b** Displacement d determination from d_{n1} and d_{n2}



determined by the measurement of the fringe order, while in Eq. 4.6 only the sum $\alpha + \beta = 2\gamma$ is known but α and β are unknown.

If the displacements are small compared to the size of the object and even more compared to the size of the ellipse, d_n can approximate the component of the effective displacement along the normal to the ellipse passing through C , i.e., lying on the bisector of the angle 2γ , equal to $d_n \sim d \cos \theta$, Fig. 4.6a.

If C is the point of the object, A the optical center of the spatial filter and B the point of observation on the photographic plate (Hologram), Fig. 4.5, the fringes are the locus of points of displacements with constant component d_n along the bisector of the angle formed by the illuminating beam and the viewing direction from any point on the hologram plate, Fig. 4.6a.

A single observation does not allow the determination of the actual displacement d in modulus and direction but only the value of d_n . The value of d can be determined by two observations made from two different points of the holographic plate. Two different views from two different points provide two different interferometric patterns. If the observation point moves continuously, also the fringes seem to move. Only the zero-order fringe that corresponds to a zero displacement in all viewing directions does not participate in this movement. The displacement vector d must have both ends belonging to two ellipses, and is therefore univocally determined.

From an analytical point of view, rather than solving the two equations for the ellipses crossing C' , the first with focuses A and B and the second with focuses A and B_2 , Eq. 4.11 is used twice for two different known values of γ :

$$\begin{cases} d_{n1} = \frac{N_1 \lambda}{2 \cos \gamma_1} \\ d_{n2} = \frac{N_2 \lambda}{2 \cos \gamma_2} \end{cases} \quad (4.12)$$

Once d_{n1} and d_{n2} in modulus, direction and orientation, d becomes uniquely known, Fig. 4.6b.

In the particular case in which the direction of d is known, only one observation (therefore a single relationship), allows the determination of the displacement. It is possible to operate in various ways:

1. Using Eq. 4.11, and the known angle δ between d and d_n , d is obtained from:

$$d = \frac{d_n}{\cos \delta} = \frac{N \lambda}{2 \cos \gamma \cos \delta} \quad (4.13)$$

2. If the object is oriented so that the displacement direction coincides with the direction of the incident light, Eq. 4.6, gives:

$$\frac{N \lambda}{1 + \cos 2\gamma} \quad (4.14)$$

4.4 Other Arrangements

More than a method for determining the displacements value, the practice suggested by [1] of drawing a family of ellipses on a large sheet of paper positioned on the holographic bench, is a way of clarifying the locations of all the elements and avoiding a long phase of pre-tests.

A series of ellipses are drawn with foci A and B , Fig. 4.5, spaced along the x axis (coinciding with the greater diameter) at a distance multiples of $\lambda/2$, in such a way that, for each ellipse, the difference between the optical paths of the object and of reference beams is always less than the coherence length, also when the point C on the object and the reference mirror are represented on the diagram by points placed on adjacent ellipses from the opposite side of the considered main ellipse.

The diagram is useful not only to help in setting up objects and optical devices on the holographic bench, but also to verify how a large body can be placed on it, remaining within the coherence length of the laser, and also for a first interpretation of the analytical relationships. The diagram also clarifies a series of possible arrangements of the optical system, Fig. 4.7.

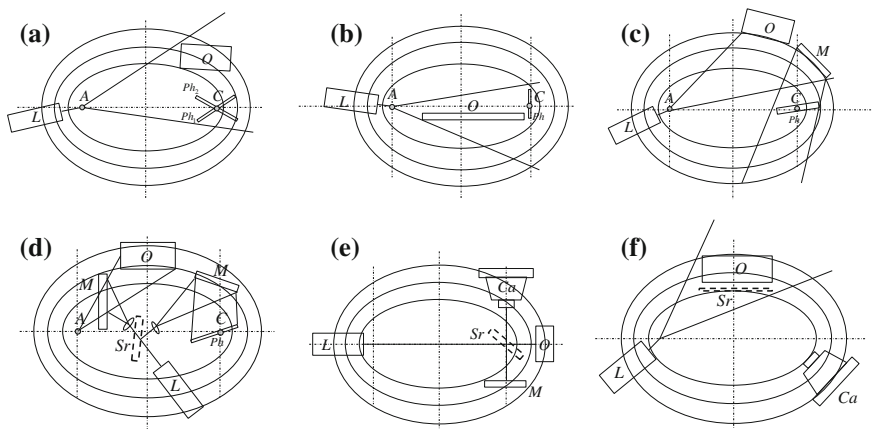


Fig. 4.7 Possible layouts of holographic benches, where L Laser, O object, Ph photographic plate, M mirror: **a** without any mirror for the reference beam, with plate in position Ph_2 gives the so-called Lippman hologram and in position Ph_1 the normal hologram; **b** suggests how to positioning a long bar; **c** is a classic hologram with the mirror for the reference beam; **d** shows an alternative set-up with reflecting mirror plus two mirrors; **e** gives a classic Michelson interferometer; finally **f** shows a new type of interferometer with semi-reflecting mirror Sr and with C photographic camera (from Amramson [1–3])

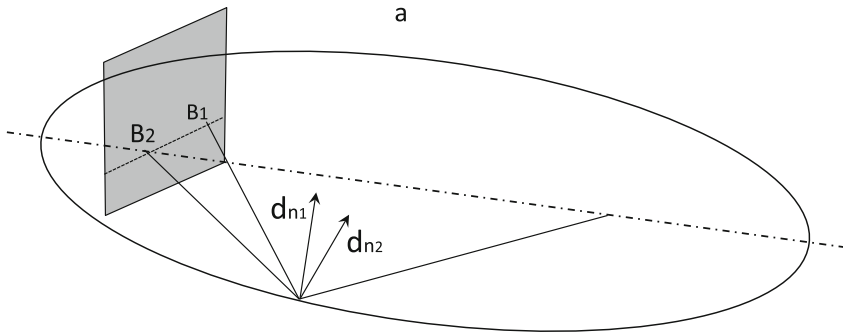


Fig. 4.8 Determination of the displacement vector lying in plane ABC

4.5 Displacement Vector Lying on Plane ACB

In the case of displacement lying on the plane ABC Fig. 4.8, two normal components d_{n1} and d_{n2} of the unknown value d are determined by two observations from the points on the holographic plate B_1 and B_2 .

4.5.1 Case Study of Displacements Determination in a Transparent Model

The application deals with the problem of determination of displacements of the surfaces of two circular discontinuities internally pressurized, embedded in a transparent block. An application to a Fracture Mechanics problem will be shown in Chap. 9. Both discontinuities are artificially generated in the middle plane of a transparent plexiglass block of $300 \times 200 \times 110 \text{ mm}^3$, Fig. 4.9. The transparent model is placed on the holographic bench as shown in Fig. 4.10.

The geometric sizes are:

- $2a = 60 \text{ mm}$ crack diameter.
- $h/a = 1.09$ distance between the centre of each crack and the symmetry plane over a .

The displacements field can be determined by the orders and distribution of the fringes. Being double exposure holograms, the integer orders are associated with the bright fringes.

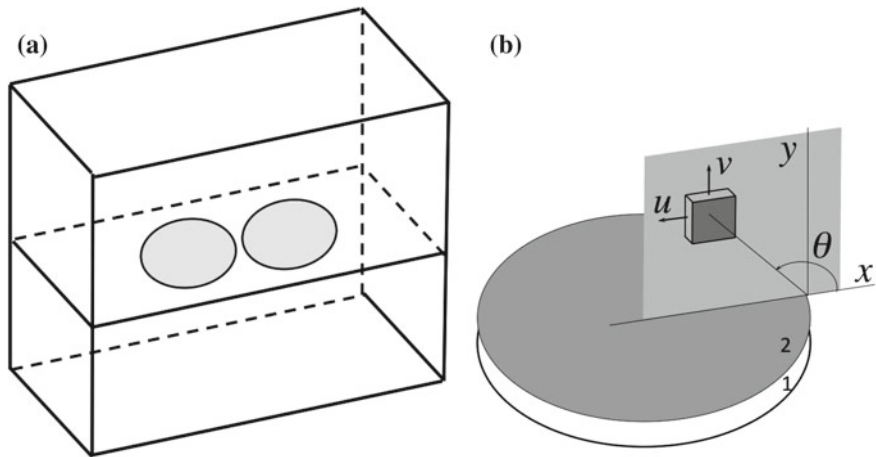


Fig. 4.9 Two co-planar circular discontinuities in a block and displacements field on a plane at a very small distance apart, parallel to their plane

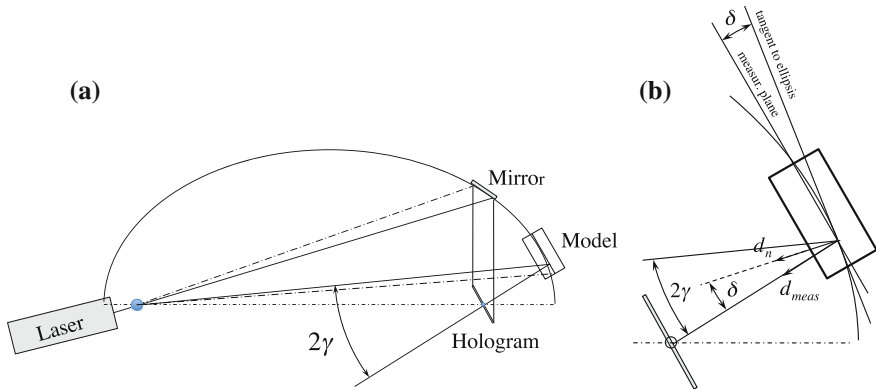


Fig. 4.10 Positioning of the transparent model on the holographic bench. **a** General set-up, and **b** detailed of displacements components

For an internal pressure of 3 bar the displacement field of one crack is shown in Fig. 4.11 and of both cracks in Fig. 4.12.

4.6 Displacement Vector Not Lying on Plane ACB

In the most general case of vector d not lying on plane ACB , two observations from two points B_1 and B_2 of the hologram plate allow only the determination of displacement components lying on the plane defined by points ACB_1 or ACB_2 .

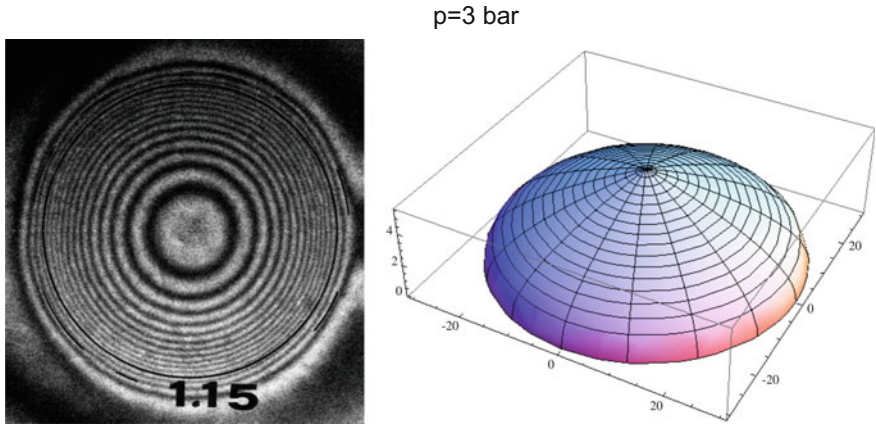


Fig. 4.11 Double Exposure Fringes of displacements field, and three-dimensional representation of one of the surfaces of two co-planar cracks in a transparent model internally pressurized at 3 bar

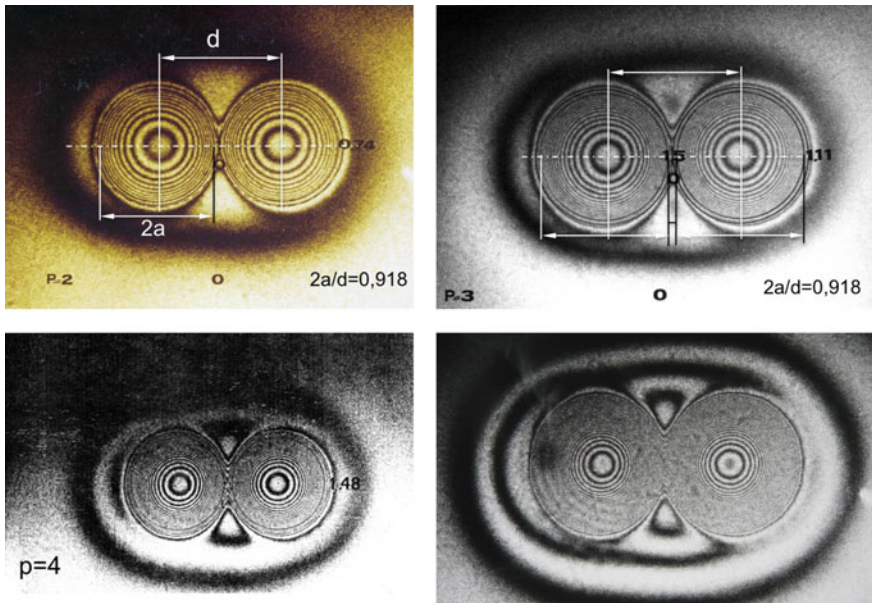


Fig. 4.12 Double-exposure Fringes of displacements field of the surfaces of two co-planar discontinuities surfaces in a transparent model internally pressurized at different pressures (2, 3, 4, 5 bar)

A third relationship Eq. 4.11 can be written with reference to an ellipse with focal points A and B_3 passing through C , where B_3 is a third observation point on the plate, not lying on plane $ACB_1 = ACB_2$, Fig. 4.13. In the same way, the relationship

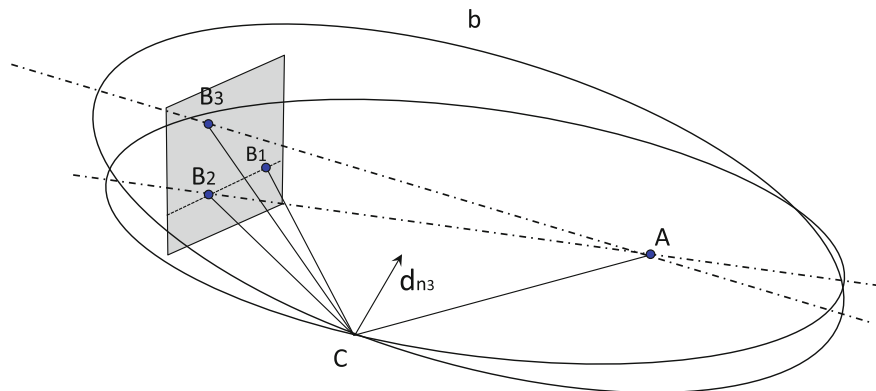


Fig. 4.13 Determination of the displacement vector not lying on plane ABC

Eq. 4.11 applies to ellipse $A B_3 C$ and a third component of the displacement d is inferred. The unknown displacement is now determined by three components d_{n1}, d_{n2}, d_{n3} in three different directions.

Vectors d_{n1}, d_{n2}, d_{n3} must be projected along the cartesian axes to determine cartesian displacement components u, v, w in directions x, y, z :

$$\begin{cases} u = d_{n1} \cos n_1 \hat{u} + d_{n2} \cos n_2 \hat{u} + d_{n3} \cos n_3 \hat{u} \\ v = d_{n1} \cos n_1 \hat{v} + d_{n2} \cos n_2 \hat{v} + d_{n3} \cos n_3 \hat{v} \\ w = d_{n1} \cos n_1 \hat{w} + d_{n2} \cos n_2 \hat{w} + d_{n3} \cos n_3 \hat{w} \end{cases} \quad (4.15)$$

or:

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \mathbf{cos} \cdot \begin{pmatrix} d_{n1} \\ d_{n2} \\ d_{n3} \end{pmatrix} = \frac{\lambda}{2} \mathbf{cos} \cdot \begin{pmatrix} N_1 / \cos \gamma_1 \\ N_2 / \cos \gamma_2 \\ N_3 / \cos \gamma_3 \end{pmatrix} \quad (4.16)$$

where symbol \mathbf{cos} means the matrix of direction cosines⁵ of the three bisectors n_1, n_2, n_3 with reference to the coordinate axes and N_1, N_2, N_3 are the fringe orders at point C for any set of three observation points 1, 2, 3 that must be non-aligned on the holographic plate.

When the plate of the hologram is small, the points are very close to one another, and the determination of the displacement is subjected to high accidental errors.

The determination of the displacement vector by more than three observations with a minimization procedure, such as the minimum square method, can be a proper solution for reducing errors.

⁵ The direction cosines of a vector are the cosines of the angles between the vector and the three coordinate axes.

4.6.1 Order Determination When No Zero Fringe Is Present

Let us consider the case of displacements contained in plane ACB .

If the unknown displacement has a known direction, from a first observation, this is, since $k_1 = 1/\cos \gamma_1$ and $k_2 = 1/\cos \gamma_2$:

$$d_x = N_1 k_1 \frac{\lambda}{2} \cdot \frac{1}{\cos \alpha_1} \quad (4.17)$$

From a second observation:

$$d_x = N_2 k_2 \frac{\lambda}{2} \cdot \frac{1}{\cos \alpha_2} \quad (4.18)$$

N_1 and N_2 are not accessible as absolute values, but their difference can be determined:

$$N_2 - N_1 = m \quad (4.19)$$

Equating 4.17 and 4.18, and taking into account the Eq. 4.19, this is:

$$N_1 k_1 \frac{\lambda}{2} \cdot \frac{1}{\cos \alpha_1} = (N_1 + m) k_2 \frac{\lambda}{2} \cdot \frac{1}{\cos \alpha_2}$$

from which the normal component of the displacement is derived:

$$N_1 = \frac{m k_2 \cos \alpha_1}{k_1 \cos \alpha_2 - k_2 \cos \alpha_1} \quad (4.20)$$

Finally, from Eq. 4.11 the normal component is given by:

$$d_{n1} = N_1 k_1 \lambda / 2 \quad (4.21)$$

and the actual displacement becomes:

$$d_x = \frac{d_{n1}}{\cos \alpha_1} \quad (4.22)$$

If α_1 and α_2 are not known, but considered equal from indirect observations, Eq. 4.20 gives an approximate expression of components d_{n1} , d_{n2} and thus of d :

$$N_1 = \frac{m k_2}{k_1 - k_2} \quad d_{n1} = \frac{m k_1 k_2}{k_1 - k_2} \cdot \frac{\lambda}{2} \quad d_{n2} = \frac{m k_2^2}{k_1 - k_2} \cdot \frac{\lambda}{2}$$

4.7 Strains from Displacements

If strains along the x axis are required, both members of Eq. 4.16 must be derived with respect to the variable x :

$$\begin{pmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial v}{\partial x} \\ \frac{\partial w}{\partial x} \end{pmatrix} = \frac{\lambda}{2} \cdot \mathbf{cos} \begin{pmatrix} \frac{\partial N_1}{\partial x} / \cos \gamma_1 \\ \frac{\partial N_2}{\partial x} / \cos \gamma_2 \\ \frac{\partial N_3}{\partial x} / \cos \gamma_3 \end{pmatrix} \quad (4.23)$$

This derivation is correct if the variations with x of the matrix terms of direction cosines and of γ cosines are neglected.

Terms $\partial N_i / \partial x$ may be approximated by finite differences $\Delta N_i / \Delta x$ obtained by interferometric images, (without knowing the absolute fringe orders), fixing a local coordinates system at the point, taking y axis lying e.g. on the surface of the body, and measuring the changes ΔN_1 , ΔN_2 , ΔN_3 between C and a second point C_0 fixed on x axis at a certain distance Δx from the point C . Similarly, the other deformation components can be determined.

The exposed methods are hard to apply, as they should be repeated point by point. It is therefore very important to do preliminary qualitative analysis of the interference fringes for identifying areas of greatest interest to which the subsequent quantitative analysis should be restricted.

Holographic interferometry offers great advantages for measuring displacements, providing:

- Total field technique suitable for laboratory use but also for industrial control departments.
- No contact measurement.
- High accuracy.
- High sensitivity.

but presents a series of disadvantages such as:

- Slow operation.
- Possible problems with rigid displacements with generation of spurious fringes.
- Not common availability of special photographic film at high-resolution (perhaps new electronic recording media with very high resolutions could come in the future, [10]). The need for high resolution film (receiver) is due to the very small fringe steps that must be recorded. The order of magnitude of the fringe step p is about the same as d_n for N equal to one: $\lambda 72 \cos \gamma$. Then for the wavelength of the light emitted by a He-Ne laser ($\lambda = 632 \text{ nm}$), the frequency of recorded lines on a photographic plate is

$f = 1/p = 2/\lambda = 2/(632 \times 10^{-6}) \simeq 3,000$ lines/mm. Special photographic plates offer such resolution and contains 10^{10} bits of information in an area of $100 \text{ mm} \cdot 125 \text{ mm}$.

Conversely, a *CCD* sensor provides an image with 10^7 bits of information or even higher, but at the present time this is still far from what is required to build a hologram.

- High stability of the optical bench (typically a sensitivity of $0.5 \mu\text{m}$ (displacement/fringe order) is required.
- High power of the laser, used in the case of the examination of large areas, is also required.
- Complexity of the optical schemes and holographic bench arrangement.
- Cost of the equipment.

References

1. Abramson N (1970) The holo-diagram II: a practical device for information retrieval in hologram interferometry. *Appl Opt* 9(1):97–101
2. Abramson N (1970) The holo-diagram III: a practical device for interpreting fringe patterns in hologram interferometry. *Appl Opt* 9(10):2311–2320
3. Abramson N (1972) The holo-diagram V: a device for practical interpreting of hologram interference fringes. *Appl Opt* 11(5):1143–1147
4. Abramson N (1981) The making and evaluation of holograms. Academic Press, London
5. Ajovalasit A (2009) Analisi sperimentale delle tensioni con la fotomeccanica. Fotoelasticita, moire, olografia, speckle, correlazione immagini, vol 1. Aracne Roma, Roma
6. Caulfield HJ (1979) Handbook of optical holography. Academic Press, New York
7. Dally JW, Riley WF (2005) Experimental stress analysis, 4th edn. McGraw-Hill Co, New York
8. Pirodda L (1974) Principi di interferometria olografica applicata alla determinazione di spostamenti e deformazioni, lezioni tenute al Corso di aggiornamento CRAD di Udine: Progettazione integr. e ass. nella Ing. Strutt
9. Rastogi PK (1999) Photomechanics. Springer, Berlin
10. Schnars U, Jüptner W (2005) Digital holography: digital hologram recording, numerical reconstruction, and related techniques. Springer, Berlin
11. Sharpe WN Jr, Sharpe WN (2008) Springer handbook of experimental solid mechanics. Springer. ISBN: 978-0-387-26883-5
12. Smith HM (1969) Principles of holography. Wiley, New York

Chapter 5

Overview of Digital Image Correlation

Luca Cristofolini

Abstract This chapter provides an introduction to the methods for measuring the field of displacement and strain on the surface of a specimen by means of digital image correlation (DIC) techniques. Both its two-dimensional and three-dimensional versions are presented. A simplified overview of the algorithms underlying such a method is provided. As DIC can lead to subtle errors, which may go unobserved, the focus of this section is on the several possible sources of random and systematic errors, and on possible approaches and practical solutions to keep such errors under control. The methods for preparing the specimen's surface are described in detail. An extensive discussion is provided about the strategies for optimizing the processing parameters and the post-processing filters in order to minimize the noise while avoiding significant loss of information. The last part of this chapter provides an overview of the volumetric version of such a method (digital volume correlation, DVC), and some applications.

5.1 Introduction

Most of experimental stress (strain) measurement techniques seen so far suffer of one of the following limitations. In some case (strain gages), accurate measurements are possible, but only at selected locations, while the remaining areas (between strain gages) are not investigated. The full-field techniques seen so far involve a significant amount of labour (photoelasticity, interferometric techniques), are associated with a limited strain resolution (photoelasticity), rely on models in place of the actual specimen (transmission photoelasticity), or are associated with significant perturbation of the actual system (reflection photoelasticity). Digital Image Correlation (DIC) is a non-contact method that enables measuring the full field of displacements and strains on the visible surface of the

test object. Analysis by means of DIC relies on a comparison between series of digital images starting from the unloaded state, and up to the selected loading conditions (possibly including intermediate steps). The surface must have a high-contrast random pattern, so that the software can univocally recognize portions of the surface (consisting of square or rectangular subsets of pixels, often called facets), and track them when the specimen is loaded, evaluating the field of displacement. From the field of displacements, the strain field on the specimen surface can then be obtained by differentiation. If images are acquired by a single camera, then planar displacements and strains (in the plane perpendicular to the direction of observation: two-dimensional case, 2D) are measured. If two cameras are used, stereoscopic vision enables obtaining the full spatial field of displacements (including out-of-plane ones: three-dimensional case, 3D) and strains on the specimen's surface. One of the strengths of DIC is that its operating principle can be adjusted to all dimensional scales, from microscopic (specimens smaller than a millimeter, with sub-micron resolution) to macroscopic (as large as several meters).

Digital Image Correlation is a grandson of the first studies on optics. DIC incorporates the theory of projective geometry and perspective vision due to Sturms and Haick, 1883; the principles of photogrammetry put forward by Finsterwald in 1899; the projective equation for stereoscopic vision solved by von Gruber in 1924; the solution of the equations for photogrammetry by Earl Church in 1945; and the matrix notation for analytical photogrammetry developed by Schmid in 1953. This combination allowed first extracting the three-dimensional shape of objects, and later analyzing shape variations and strains by least squares solution of correlation problems.

When digital photography became commercially available, in the Sixties and Seventies, it was first exploited to enable spatial vision and recognition of objects in the field of robotic and artificial intelligence. At that time, photogrammetry in engineering was mainly seen as a possible application of the enormous potentiality of coherent light provided by the recently invented laser: the focus in those years was on holographic and speckle laser interferometry, and on moire methods.

The concept of modern Digital Image Correlation dates back to the early Eighties, when its 2D implementation was explored at the University of South Carolina as a means of measuring in-plane strains, strain gradients, and crack initiation and propagation in brittle materials. One of the first studies where digital images were exploited to measure surface strains was published in 1982 [1]. At that time, a pure planar technique was proposed, where a single camera enabled measuring in-plane (but not out-of-plane) displacements and strains.

The first three-dimensional applications followed, in the early Nineties, when the principles of stereoscopic vision were combined with those of planar digital image correlation [2]. The system, which incorporated two digital cameras, enabled measuring the three-dimensional state of strain around a notch. The same approach was also applied to the investigation of the three-dimensional motion of a fluid with

high-contrast particles in suspension (Particle Movement Velocimetry, PIV). In the late Nineties applications to at the micrometer and nanometer scale were found.

Further improvements included the effect of perspective on the detected shape of the subsets of pixels (facets). Epipolar constraints were also incorporated in the analysis, to improve the accuracy by taking into account the correspondence of geometric elements between two conjugated images. A better image reconstruction process [3] further enhanced the calculation of displacements: high-order splines were introduced to interpolate the pattern images. This enabled improving the accuracy of displacement measurements, down to the order of 0.01 pixel (therefore, the physical accuracy depends on the magnification factor, but it can easily be demonstrated that this corresponds to sub-micron accuracy when a camera with millions of pixels is used on a field of view of some centimeters). At this point, the correction of optical distortions became extremely important.

The practical feasibility and its spatial resolution improved over time as: (i) digital cameras improved (a resolution of at least one million of pixels is typically necessary to obtain reasonable resolution on a suitable area); (ii) computation power increased (the algorithms for image correlation can be quite cumbersome when high-resolution images are processed); more accurate and efficient numerical algorithms were developed.

Digital Image Correlation as a modern strain analysis technique was outlined in the classic handbook of Dally and Riley [4]. More recently, an entire volume was dedicated to the theory and methods of image correlation [5].

5.2 Operating Principle

The principle of Digital Image Correlation relies on sets of images of the surface of the test object in the original state, and in a displaced and/or deformed state. A *sensor* (which can either be the single digital camera, in the 2D implementation, or a calibrated set of two cameras, in the 3D implementation) collects a sequence of images as the test object is displaced and/or deformed. A software then computes the displacements at any point on the visible surface of the test object. This enables obtaining the field of displacements and, by derivation, the field of strain. A key problem of DIC is that strain patterns tend to be affected by large noise. In fact, derivation acts as the opposite of a filter, and enhances the noise in the displacement field computed from noisy digital images [6].

To enable the software to univocally identify each point on the specimen's surface, and track them across the subsequent images acquired while the specimen is displaced and/or strained, the specimen's surface must have a high-contrast random speckle pattern, Fig. 5.1. It has been demonstrated that a suitable pattern facilitates the interpolation of the distribution of light intensity in the acquired images, and enables estimation of displacements with sub-pixel accuracy [3].

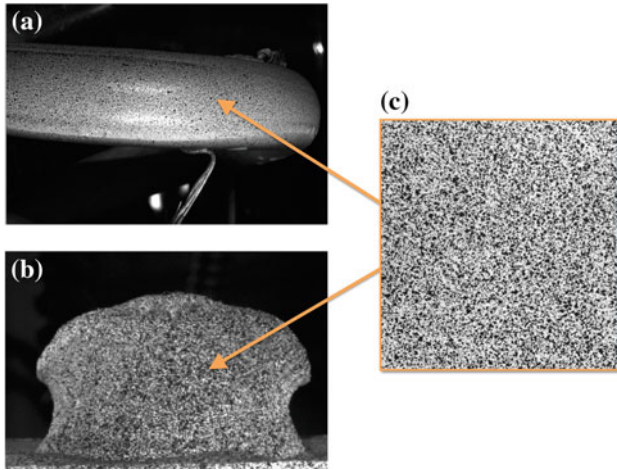


Fig. 5.1 Examples of specimen with a random *black-and-white* pattern: **a** coil of a spring; **b** anterior view of a vertebra; **c** detail of speckle pattern

While in the following pages the description will focus on a comparison between two states (undeformed and deformed), DIC can be extended to compare three or more steps. This is particularly useful in the case of large displacements/strains, and in general, when the progress of strain over time must be investigated.

5.3 Computation Area (Facet)

It is useless to search for a correspondence of a single pixel in two images. In fact, the grayscale value of a single pixel can be found at many of other pixels in the second image. The solution is to search for the correspondence of a small neighborhood around the pixel of interest. The acquired images are then divided in smaller subsets of pixels (facet), which are the elementary computation area. Each facet is associated with one point for which displacements are computed (typically the center of the facet itself). Therefore, the larger the number of facets, the higher the computation accuracy.

Usually, squared facets of $N \times N$ pixels are used, where N represents the facet size. In some specific cases (large strain in one preferential direction; specimens anisotropy), rectangular facets may be preferable. The size of the facet to some extent can be seen as something similar to the size of the grid of a strain gage. The advantages of larger facets are: robust identification and better correlation in subsequent images; more accurate measurement of displacement and strain; and less noise in the computed displacements and strains. At the same time, large facets may result in loss of the information associated with high strain gradients. In addition, the boundaries of the region of interest cannot be solved with a high accuracy: the extension of the affected region is directly related to the facet size. The computational

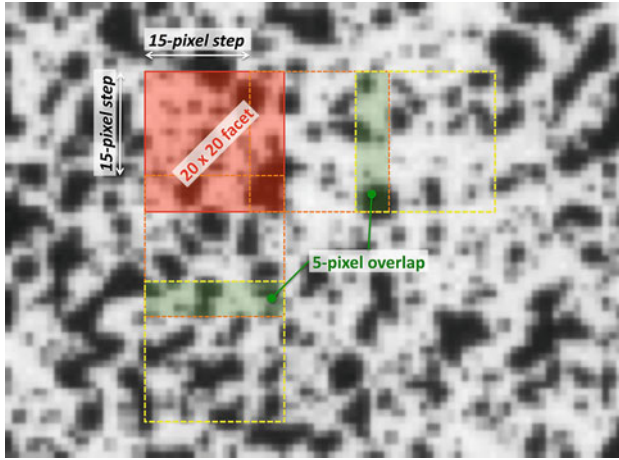


Fig. 5.2 Detail of a portion of the specimens' surface with a random *black-and-white* pattern: two facets are indicated (in this case they are squared, of $N \times N = 20 \times 20$ pixels, with a partial overlap of $M = 5$ pixels, and a step of 15 pixels between facets

cost increases proportionally with the extension of the computation area (i.e. with the square of the facet size, N^2). Smaller facets enable observing local effects and gradients. The disadvantages of smaller facets are: noisier calculation of the strain pattern, and high risk of lack of correlation in some domain.

To provide some redundancy, facets may be partially overlapped by a certain number of pixels, M Fig. 5.2. The step between consecutive facets determines the amount of overlap between facets. An overlap to a certain extent allows a more robust estimate of the displacements (at a higher computational cost) when the original input images are affected by noise. When no overlap is applied ($M = 0$ pixels), the facets are just tiled one next to the other. The maximum possible overlap is close to the actual facet size ($M = N - 1$ pixels). In this case, the maximum possible number of data is computed for a given image resolution. This clearly results in a higher computational cost.

In principle, consecutive facets could even be juxtaposed with a gap between each other (i.e. with a negative overlap M). However, this would result in a loss of information and of spatial resolution, with limited advantage in terms of computational cost.

5.4 The Speckle Pattern

In order for the DIC method to work effectively and accurately, the surface of the specimen (at least the portion under observation) must present a specific pattern. In fact, as will be explained better in the following pages, image correlation exploits non-uniform light reflection by the surface under observation. A uniform surface

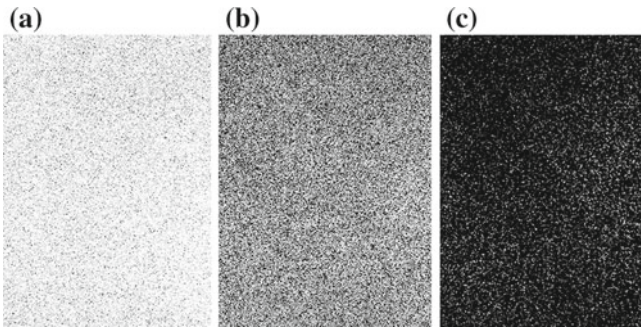


Fig. 5.3 Examples of speckle pattern: **a** too light (the *black dots* cover far less than 50% of the surface); **b** optimal (*black* and *white* balanced); **c** too dark (the *black dots* cover far more than 50% of the surface). Courtesy of Marco Palanca (University of Bologna)

would offer no features nor light gradients to be tracked. Therefore, a first requirement is to have a pattern on the specimen's surface. To allow the software to univocally identify the regions on the specimen's surface, such a pattern must be random, so that no area resembles any other. Such a pattern needs to be an integral part of the specimen under observation (i.e. it must move and strain together with the underlying material).

In most cases, the test specimen does not present a natural speckle pattern. For this reason, a speckle pattern is often painted on the specimen's surface by means of an airbrush gun. As a high-contrast speckle pattern is desirable to reduce the effect of noise, in most cases a black-on-white pattern is prepared, although a white-on-black one would be suitable as well. Most DIC systems exploit black-and-white cameras (each pixel is associated with a grayscale value), which allow achieving sharper images. An alternative to airbrush spraying consists of application of toner powder. The advantage is the possibility of controlling the size of the applied particles. The main disadvantage of toner powder is that in this case the black dots displace, but do not strain together with the underlying material. Another option is that of modifying the material surface by etching, or by photolithography. If the specimen material allows (as for some biological tissue), the material itself can be dyed (e.g. with methylene blue), and subsequently prepared with a white pattern [7]. In all cases, to achieve optimal results, it is important that the fraction of area covered by the black (or white) speckles is approximately the same as the area where the white (or black) background is visible Fig. 5.3.

Digital image correlation in principle can be applied to an object of any size, as long it can be viewed by the cameras provided with suitable lenses. If a camera sensor with a given resolution is used, each pixel of the image will correspond to a certain size on the physical object, depending on the magnification factor. A pixel will correspond to a larger area if a large specimen is observed from a distance

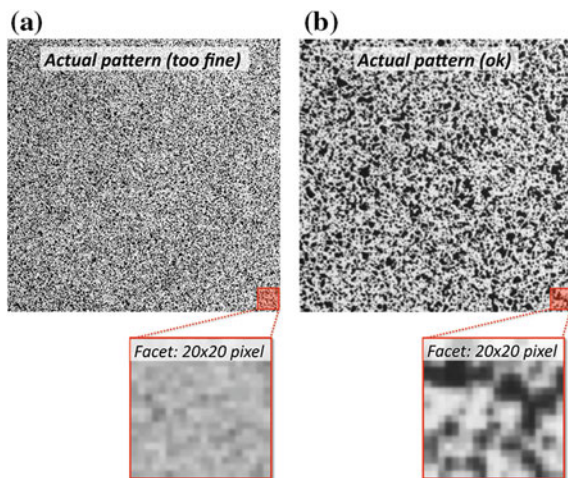
(or with lenses with a wide angle), whereas the same pixel will correspond to a small area if a small specimen is observed from close (or with lenses with a narrow angle).

Strict requirements apply to the optimal speckle size [5, 8–10], Fig. 5.4. If a speckle is smaller than a pixel, it cannot be properly detected: the pixel where such a speckle appears will have a tonality of grey, depending on the amount of black and white included. This goes against the need for a high-contrast speckle pattern. To over-sample the features in each speckle pattern, the minimum speckle size should be between 3 and 5 pixels.

Conversely, if the speckles are larger than the minimum recommended size, they also require using larger facets (i.e. sizing more pixels). In fact, in order for each facet to be unique, the facets must be sufficiently large to include a significant number of features (speckles). Therefore, excessively large speckles (i.e. exceeding 5 pixels in size) require unnecessarily large facets (i.e. containing a large number of pixels) to track them. This in the end reduces the spatial resolution of the method, and the suitability of the DIC to identify local gradients. In summary, an optimal pattern has a minimum speckle size between 3 and 5 pixels (to be suitably captured by the camera), and a low scatter (to avoid uselessly large speckles). At this point, it should also be clear why high-resolution cameras can help to improve the resolution and accuracy of DIC applications.

There may be cases where the surface naturally presents the required features, but in most cases such a pattern needs to be prepared on the surface. While DIC is considered a non-invasive measurement technique, DIC observation still somewhat disturbs the system under observation (following Heisenberg uncertainty principle): when a pattern is applied to the specimen, this constitutes a mechanical reinforcement

Fig. 5.4 Examples of speckle pattern on the specimen's surface: **a** if the speckles are too small, this results in a blurry gray distribution among pixels; **b** ideally each speckle should have the size of few pixels



(in most cases this is absolutely negligible). In all cases, the specimen is to some extent disturbed by the high-intensity light used to illuminate the area under observation. Furthermore, care must be taken to avoid that the paints used modify the specimen's material properties (mainly because of the presence of solvents).

5.4.1 Case Study: Optimization of the Speckle Pattern

The quality of strain measurements by digital image correlation (DIC) strongly depends on the quality of the pattern on the specimen's surface. An ideal pattern should (i) have an average size that exceeds the image pixel size by a factor of 3 to 5 (so as to have optimal pixel and facet size, improving resolution and correlation); (ii) have a limited scatter in terms of speckle sizes (so that most speckles are indeed of the desired size). In most cases, speckle patterns are prepared spraying black paint with an airbrush gun so as to partially cover a previously prepared white background. Here we will show how to define the ideal speckle size in relation to the specimen size and acquisition system. Furthermore, we will provide practical guidelines to identify the optimal settings of an airbrush gun, in order to produce a pattern that is as close as possible to the desired one while minimizing the scatter of speckle sizes. Patterns of different sizes were produced using an airbrush gun with different settings of the four most influential factors (dilution, airflow setting, spraying distance, and air pressure) [8]. A full-factorial DOE strategy was implemented to explore the four factors at two levels each.

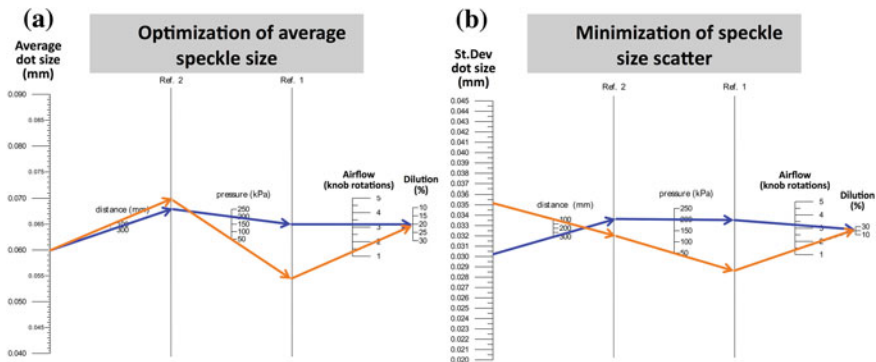


Fig. 5.5 Example of nomograms to calculate the optimal settings for a given airbrush. Based on the size of the area under investigation, and on the resolution of the camera, the typical pixel size can be computed. This enables computing the desirable average speckle size can be estimated. Nomogram (a) enables identifying the combination of settings to obtain the desired average speckle size. Among such combinations, the settings that provide a reduced variability can be selected using nomogram (b). The dotted lines are examples of two possible solutions to obtain the same average speckle size (0.060 mm), with a standard deviation of 0.030 and 0.035 mm respectively. Courtesy of Giacomo Lionello (Rizzoli Orthopaedic Institute, Bologna, Italy)

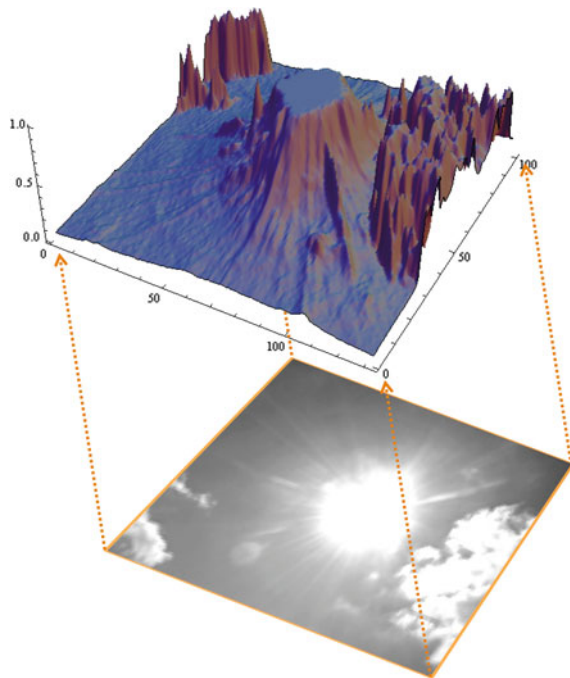
The distribution of sizes of the speckle dots was analyzed assuming a truncated Gaussian distribution: the average speckle size was calculated, together, with the standard deviation. A mathematical model was built to enable prediction of the average speckle size in relation to the airbrush gun settings. It was shown that optimal adjustment of the airbrush allowed producing a pattern with the intended average speckle size, and a limited scatter of speckle sizes. This enables matching the ideal distribution of speckle sizes for DIC, Fig. 5.5.

5.5 Operating Principle of 2D Image Correlation

Although state of the art DIC usually is implemented in its three-dimensional version (i.e. featuring two cameras, which allow the full spatial reconstruction of displacements and strains), the operating principle of DIC is easier to understand starting from its two-dimensional embodiment. The 2D version of DIC exploits a single camera, and enables measuring the displacements and strains in a plane that is perpendicular to the direction of observation.

As we said earlier, each facet includes a finite subset of pixels. Each pixel is associated with a given grayscale value (in most cases, color images offer no advantage). If the camera operates on an 8-bit basis, each pixel is assigned a value between 0 and 255. Therefore, each facet is described by an average light intensity, and by a

Fig. 5.6 Example of grayscale image with a non-uniform intensity. The portion with clouds (right side of the picture) resembles the appearance of the random speckle pattern that is desirable for DIC: the distribution of grayscale values contains a unique signature if the surface presents a random speckle pattern. Conversely, other regions are unsuitable either because the signal saturates (near the sun), or because they lack of features and gradients (uniform region in the sky)



two-dimensional internal variation, Fig. 5.6. Such an average grayscale value and its variability are hopefully unique to each facet thanks to the random nature of the speckle pattern on the specimen’s surface. This allows univocally identifying and tracking each portion of the specimen’s surface across the frames.

5.5.1 Displacements and Strains

When the specimen is loaded, each point on its surface may undergo a certain displacement. In the case of a rigid body motion (i.e. with no strain), all the points will be subjected to the same displacement vector. Conversely, if the specimen undergoes a generic combination of displacement and deformation, each point on its surface may be subjected to different displacement vectors. Consequently, a segment on the specimen’s surface will undergo a state of displacement and strain, Fig. 5.7. The strain magnitude can be obtained by deriving the components of displacement (the second order terms need to be included in the case of large deformations). If we consider a facet, a similar concept will apply: the facet will rotate and/or translate with no deformation in the case of rigid body motion. Conversely, when the specimen undergoes a combination of displacement and strain, the facet itself will deform, while moving, Fig. 5.8.

Due to the speckle pattern, each point on the specimen’s surface has its own black/white tonality. Therefore, each point and each pixel in the facet is associated with a light intensity (recorded in terms of grayscale value). We can assume that the displacement and strain imposed to the specimen has not altered the light intensity at each of its point. This holds true if: (i) the state of strain is such as not to compromise

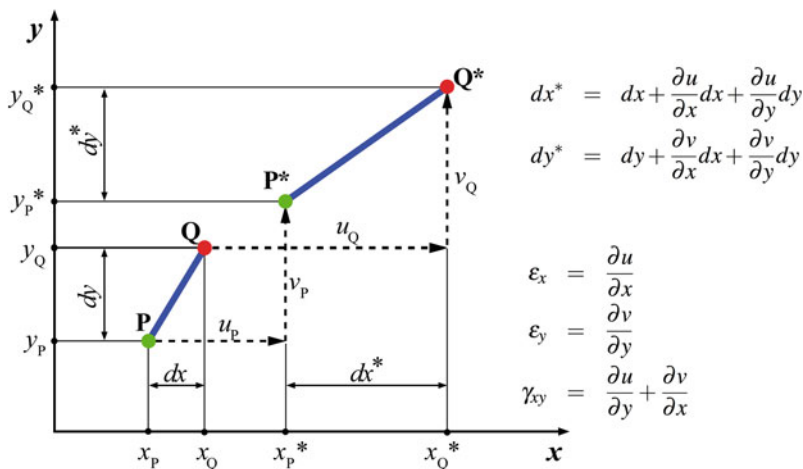


Fig. 5.7 Displacement and deformation of a segment on the specimen’s surface

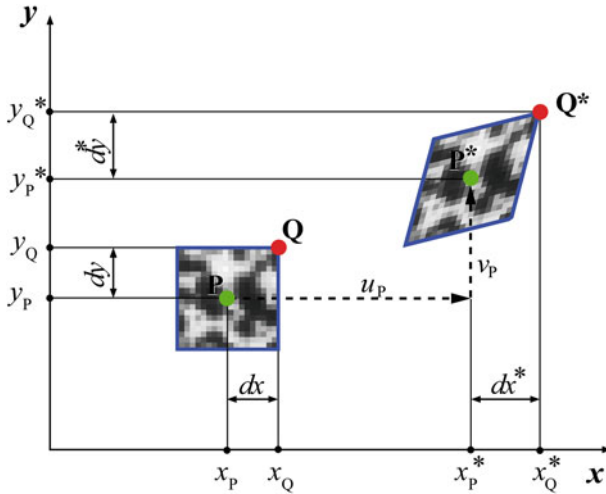


Fig. 5.8 Displacement and deformation of a portion of the specimen's surface (facet). If the specimen is uniformly illuminated, the light intensity reflected by each point of the surface (e.g. P, Q) is the same before (I(P), I(Q), etc.) and after displacement/deformation (I(P*), I(Q*), etc.)

the surface appearance (e.g. no matting due to yielding), and (ii) a uniform lighting it provided. In the case of large strains, each point may displace by a large number of pixels, and this can be detected rather easily. In the case of small strains, displacements can be smaller than a pixel. In this case, an interpolation of the grayscale values enable sub-pixel resolution in measuring displacements, Fig. 5.9. For instance, a bilinear interpolator or a bicubic spline can be used to interpolate the light intensity for a number of neighboring pixels. It has been shown that a bilinear interpolation can cause a systematic error of up to 50% on the computed strain. A bicubic spline can reduce such an error to few percents (at a higher computational cost), providing sub-pixel resolution in measuring displacements (of the order of 0.01 pixels) [3].

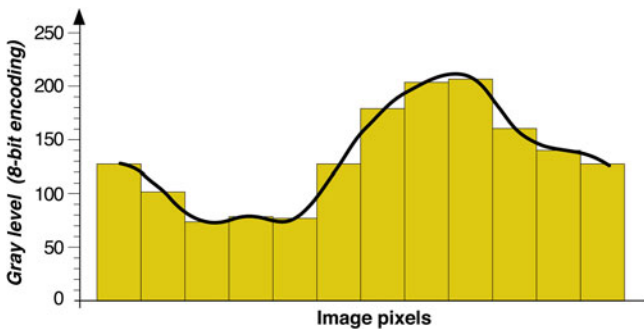


Fig. 5.9 Interpolation of the grayscale values along a line by means of a cubic spline: in DIC, the spline parameters are used to summarize the grayscale distribution in each facet

If, for instance, a pixel corresponds to 0.1 m on the specimen's surface, this means that displacements can be measured with a precision of 1 μm .

The correlation between two (or more) subsequent images is based on the division of the images into facets. Once the light intensity distribution in the undeformed state has been measured (and *summarized* by computing the interpolating functions), the facet that better resembles (in terms of interpolating function) the undeformed one can be sought in the deformed images by minimizing a suitable cost function that measures the cross-correlation between two images [11]. It must be noted that the result of a cross-correlation between images depends on the actual similarity/variations of the observed object, but also on variations over time (i.e. between acquired images) of the light intensity. Therefore, it is extremely important to illuminate the specimen with a constant light (fluorescent lights must be avoided due to their intrinsic flickering). Furthermore, normalization methods must be adopted to minimize such effects.

The details of the algorithms used to track the facets are the core of the DIC method, and vary between software packages. The easiest way to illustrate the correlation process is to follow an iterative matching, Fig. 5.10: as a first step, a rigid translation of the facet is assumed (initially neglecting the strain components, i.e. the derivatives of the displacement field), so as to track the barycenter of the facet. In a second stage, the normal strains in both directions are taken into account (i.e. by considering the first derivatives of the displacement field). Finally, also the shear strains are considered (i.e. the cross-derivatives of the displacement field). The procedure can be iterated as long as the cost function exceeds a certain tolerance value.

The algorithms most frequently implemented in the DIC software are variations of the Levenberg Marquardt non-linear optimization (this is effective when some variables have low gradients while others have high gradients): the straight-forward gradient steepest descent algorithm; the Hessian matrix approximation; the Newton Raphson (either with a local quadratic functional form, or in combination with a steepest descent).

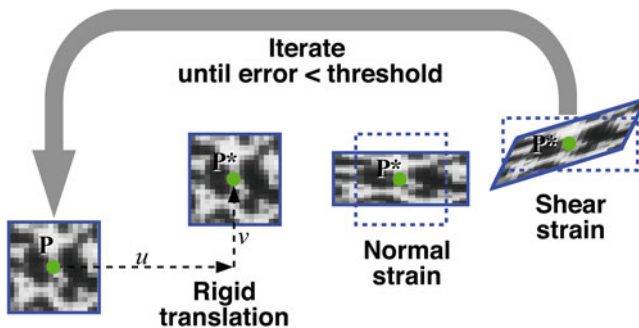


Fig. 5.10 Exemplification of the steps of an iterative matching procedure to track a facet

5.6 3D Image Correlation

A DIC system consisting of two cameras allows measuring the displacements in three dimensions (including those out of plane, in the direction of the cameras). This enables measuring the spatial state of strain, which is particularly advantageous for specimens with a complex geometry.

5.6.1 Stereoscopic Vision

The first step of 3D image correlation consists of matching the images from two cameras, exploiting stereoscopic vision to obtain a three-dimensional description of the specimen's visible surface, including the grayscale values associated with the pattern present on its surface. In fact, while a single camera allows measuring the displacements only in a plane perpendicular to the direction of observation, displacements towards the camera (or away from the camera) can be detected by comparing images from two cameras. In the case of dynamic tests, it is extremely important that frames are synchronously acquired by both cameras (this constraint is less important if static measurements are involved). In order for the stereoscopic vision to be effective, the distance between the two cameras must be sufficient, so that the directions of observation for a same point from an angle of at least $20^\circ \div 25^\circ$. While for narrower angles a trigonometric solution is still theoretically available, computation of out-of-plane displacements is less accurate as the problem is ill-posed.

Knowing the relative position and alignment of the two cameras, the position in space of any point of the test object visible by the two cameras can be computed by triangulation based on the coordinates of such point in the images acquired by the respective cameras. The triangulation process requires a calibration of the hardware: the system is trained by acquiring images of known, highly precise and stable calibration targets, Fig. 5.11. The calibration procedure slightly varies depending on the software. In all cases, it consists of acquiring images of the calibration target with different poses (i.e. translated and rotated) within a calibration volume. The software computes the intrinsic system parameters so as to minimize a cost function. Such calibration takes into account a number of factors such as: position and alignment of the cameras; optical magnification factors; in most cases optical distortions.

It must be clear that only the portion of the specimen that is viewed by both cameras can be investigated. While almost identical fields of view can be obtained for flat specimens, this is not the case for curved specimens, where the cone of view of each camera encompasses a portion of the specimen: as the cameras are at different points, such portions only partly overlap.

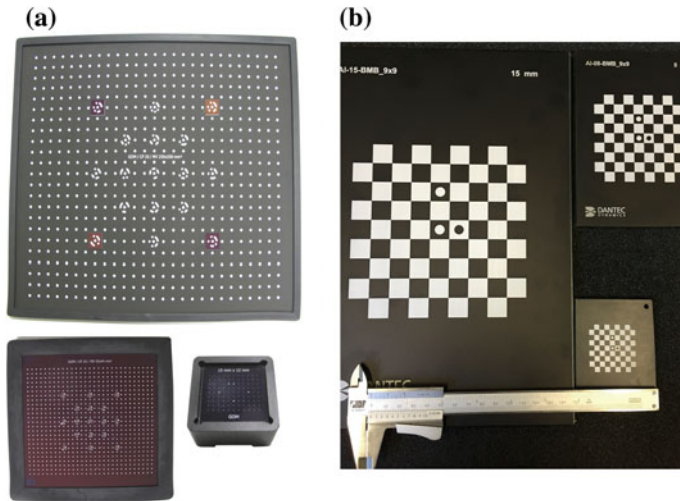


Fig. 5.11 Examples of calibration targets of different sizes: **a** adopted by GOM mbH (Braunschweig, Germany): the largest one sizes 250 mm by 200 mm, the smallest one 115 mm by 12 mm (pictures not to scale); and **b** adopted by Dantec Dynamics A/S (Skovlunde, Denmark): the largest one sizes 150 mm by 150 mm, the smallest one 35 mm by 35 mm, Pictures courtesy of Giacomo Lionello (Rizzoli Orthopaedic Institute, Bologna, Italy) and Marco Palanca (University of Bologna, Italy)

5.6.2 Rectification

If the stereo matching is performed for each point of the image with no particular strategy, this can be very slow. For a calibrated stereo system, an optimized approach can be used, called rectification (a common method in computer vision). Based on the calibration function of the cameras-lenses system, the paired images can be transformed (homographic transformations) so that all epipolar lines become horizontal. Therefore, each transformed point corresponding to a given point of the specimens' surface is found in the same scan line of both matched stereo images. This makes it much faster to *interrogate* the paired images for corresponding points, as after rectification the corresponding points are found with the same coordinate in both stereo images. The use of rectified images makes it possible to use simpler (i.e. computationally lighter) shape functions in the subsequent steps. An additional advantage of rectification is that it allows correcting errors due to optical distortion of the camera lenses. In fact, in the description of an optical system composed by thin Gaussian lenses, it is assumed that each lens is perfectly aligned. However, small malpositioning of the lenses cannot be completely avoided due to technological problems. In the rectification phase, such errors can be to a large extent compensated, and the rectified images correspond all to the same focal length.

5.6.3 Stereoscopic Matching

The stereoscopic matching is the following phase of processing paired stereo images. It consists in the identification of the corresponding points in the paired images. In this phase, the image *depth* is calculated, i.e. the distance of each point with respect to the camera system.

There are several strategies to address this problem. Two families of algorithms can be found in the literature. The first family consists of feature-based strategies. This type of algorithms aims to identify features that can be easily identified and traced, such as lines, segments, and corners. Such algorithms can be very efficient. However, they do not allow building a full-field correspondence, but only limited to the features recognized in the image. The second family relies on dense-stereo strategies, allowing to build a full-field disparity map covering the entire observed surface. The dense stereo strategies can be further divided into local, global, and semi-global. They all consist in minimizing a certain cost function; the difference lies in the optimization strategy. The global-type algorithms provide the best results as to some extent they incorporate some continuity assumption which allow reducing the noise. However, they are associated with the highest computational cost and currently are seldom implemented in commercial software. Local algorithms conversely rely on a local support, the correlation window, which allows improving the signal-to-noise ratio. The main limitation of local algorithms is that they do not necessarily provide a completely populated disparity map, as in some regions the algorithm might be unable to solve the optimal problem. In this case, de-correlated areas exist within the region under investigation. In such de-correlated areas, further calculations (of displacements and strains) are not possible.

5.6.4 3D Correlation

Once the full three-dimensional surface of the object has been built from stereoscopic matching, a three-dimensional description of the surface is available, where each point is associated with its grayscale value. At this stage, the correlation algorithms described above for the planar problem can be applied.

5.7 Typical Arrangement of a DIC System

As the quality of the image (lack of distortion, limited noise, thermal stability, etc.) is extremely important, the hardware components for a DIC system must be chosen with extreme care.

The sensor of a digital camera can either be a CCD (charge-coupled device) or a CMOS (complementary metal-oxide semiconductor). CCD sensors create high-quality, low-noise images. CMOS sensors are more susceptible to noise. Because of

the need of high-quality, low-noise images, CCD sensors are preferred in DIC application, although they are more expensive.

In a CCD, the light reaching each pixel of the sensor is converted in an electrical charge. The next step is to read the value (accumulated charge) of each cell in the image. In a CCD device, the charge is transported across the chip and read at one corner of the array. An analog-to-digital converter turns each pixel readout into a digital value.

The resolution of the camera sensor (measured in millions of pixels) determines the overall resolution capability of the system. If a higher resolution is adopted, this corresponds to smaller pixels when the same physical area is observed. Alternatively, if one wishes to keep the same pixel size, to observe a larger area a higher number of pixels is required. In the end, a higher camera resolution enables observing a larger specimen, with better spatial resolution.

The frame rate of the camera determines what type of phenomena can be observed. Standard cameras usually provide a frame rate between 10 and 50 frames per second, and are suitable for static or slow-varying conditions. High-speed cameras (up to 100,000 frames per second) are indicated for very fast events (destructive tests, impacts, etc.).

Each camera must be provided with a suitable lens. If two cameras are used (3D DIC), they must be equipped with identical lenses. To achieve high-quality images

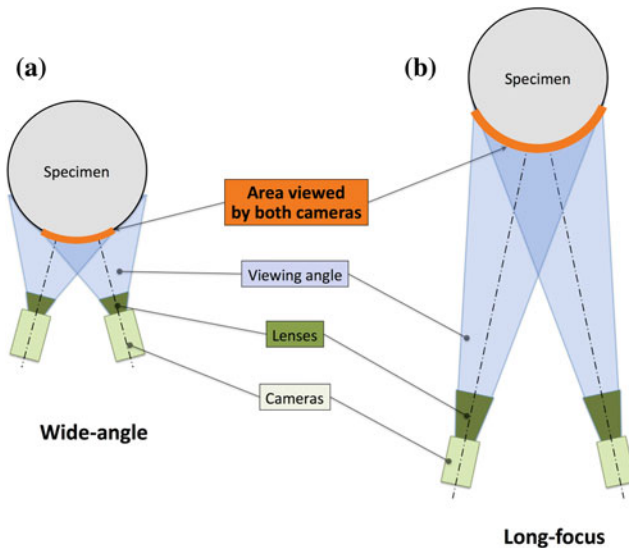


Fig. 5.12 The portion of test specimen that can be viewed by both cameras of a 3D DIC system depends on the distance of the DIC from the specimen, and on the focal length of the lens. **a** Wide-angle lenses can be placed closer to the specimen. This results in a smaller portion of the specimen seen by both cameras, due to occlusion. **b** Long-focus lenses need to be placed farther from the specimen. As a result, the two cameras see a larger portion of the object

with minimal distortion, top-quality lenses need to be used. To support a wide range of applications, the lenses must support a wide range of apertures (f-numbers). A low f-number (high aperture) is needed if lighting is poor; this is associated with limited image sharpness and limited depth of focus. High f-numbers (low aperture) is recommended when possible (sufficient light) to achieve sharper images and better depth of focus.

Using a wide-angle lens enables observing a larger specimen without the need of moving the camera far away. However, this is associated with a higher image distortion, which can result in higher errors in the subsequent computation. Furthermore, as a wide-angle lens enables keeping the DIC system closer to the specimen, this can result in a smaller area to be visible by both cameras in the case of a three dimensional specimen, Fig. 5.12a). Conversely, long-focus lenses must be placed at a larger distance from the test specimen to view same area as with wide-angle lenses. In this case, the portion of the specimen seen by both cameras is larger, resulting in a larger area being investigated, Fig. 5.12b). Telecentric lenses are a completely different class of lenses. These lenses provide a minimal optical distortion, as telecentric lenses collect an image of an object having the same size as the lens itself. The main limitation of telecentric lenses, apart from their cost, is that they must be at least as large as the observed object, and therefore are unsuitable (nor even available) for very large test specimens.

The portion of specimen investigated by the DIC with two cameras is limited to the region that can be viewed at the same time by the two cameras. In all cases this can never be more than 180° around the specimen. Although this is not very common, a DIC system can be extended by using three or more cameras, to expand the portion of specimen included in the analysis. In principle, the number of cameras could be increased as much as needed to surround (and investigate) the entire test specimen.

The DIC system must also include dedicated light sources. A first requirement of DIC light sources is that of stable illumination, as fluctuations and flickering causes unacceptable differences between subsequent frames. The preferred light source is based on high-power LED. Both white LEDs (the most common ones) and green ones (which operate in a range of frequency that is better detected by black and white CCD sensors) can be used. The light intensity to some extent determines the quality of the image, and as a consequence, the final error in the computed strains and displacements. In fact, if light is insufficient the operator has to use large aperture (small f-numbers, which results in less sharp images and shorter focal depth), or long shutter time (which can result in motion blur, especially during dynamic tests). Conversely, generous illumination allows obtaining sharp images (high f-numbers) together with lack of motion blur (thanks to short shutter times).

Several all-in-one DIC systems are available on the market, supplied by manufactures such as Dantec Dynamics A/S (Skovlunde, Denmark), GOM mbH

(Braunschweig, Germany), LaVision (Goettingen, Germany), Correlated Solution (Columbia, SC, USA).

The software for processing the images is generally provided together with commercial system. Such software usually handles the entire process: calibration of the system, image acquisition, stereoscopic matching, image correlation, calculation of displacement and strain field, filtering and post-processing. In addition, several other software packages, such as Matlab (MathWorks, Natick, MA, USA) or Mathematica (Wolfram, Champaign, IL, USA) incorporate a library for image correlation.

5.8 Theoretical and Practical Problems with DIC

Although DIC is sometimes perceived (and sold) as a turnkey system, it should be used with extreme caution because DIC measurements can be affected by very large (and subtle) problems and errors. The first issue is lack of correlation: while this can be a problem (part of the specimen cannot be investigated), it is always detectable and, to some extent, possible to fix. The second issue relates to the quality of measurement itself. Displacements are generally measured with excellent accuracy (small systematic error) and precision (small random error), well below the pixel size. Strain measurements can be quite accurate (very small systematic error), but are generally quite imprecise (very large random error). Any DIC system, either commercial or homemade, needs to be extensively tested to quantify the errors affecting displacement and strain measurements, and to identify the optimal hardware and software settings for the specific application.

5.8.1 Possible Problems Associated with Stereoscopic Matching

The third dimension is associated with some specific problems related to image correlation. In fact, while recognizing the same facet at different instants in time (i.e. in consecutive frames) involves a simple affine transformation in the two-dimensional case, the same operation becomes more complex in three-dimensional DIC.

A second problem is related to the angle between the optical axes of the two cameras. Increasing such an angle allows estimating more accurately distances and displacements in the third dimension (i.e. towards and away from the cameras) as the trigonometric problem is better conditioned. However, a larger angle is also associated with an increased risk of lack of correlation in some areas, both because of the additional distortion caused by prospective vision, and because of some portions of the object being hidden, for at least one of the two cameras (occlusion, Fig. 5.12).

5.8.2 *Observation Angle*

A limitation to the number of facets that are correctly recognized and correlated derives not only from the need of both cameras to view the same area, but also from the angle of observation of the individual cameras. When performing 2D DIC it is obvious to place the camera right in front of the area under investigation, so that the camera can best detect in-plane displacements and strains. In fact, if the axis of the lens is not orthogonal to the surface of the test specimen, the estimation of displacement and strain is biased by a systematic error due to perspective distortion.

In the 3D case (with two cameras), the perspective distortion is taken into account during the stereoscopic matching. However, other problems may still arise. In fact, when the pattern is prepared by spraying the black speckle dots, they form micro-asperities on the specimen's surface, so that when the optical axis of the camera is almost tangent to the specimen's surface, this will appear extremely dark as mainly the dark spots (which protrude on top of the white surface) will be seen. Therefore, if the specimen's surface is curved, some areas can be seen correctly (where the optical axis is nearly perpendicular), but the boundaries of such regions (optical axis nearly tangent to the surface) may appear very dark, and correlation in such areas may fail. A partial solution of this problem consists in taking more images of the specimen by moving the DIC system around it, if the test allows doing so (i.e. if a non-destructive phenomenon is being investigated). Another (more expensive) solution consists of using three or more cameras to view the specimen from different angles.

5.8.3 *Noise of the Input Digital Images*

One of the main sources of error in digital image correlation is the noise affecting the input images, which propagates to the estimated displacement and strain fields. Any electronic measurement is affected to some extent by systematic error and random noise. Image acquisition by means of digital camera makes no exception. In fact, the readout from each sub-sensor of a CCD (i.e. each pixel) is affected by both types of error. Therefore, error appears as a noise in the images acquired by the CCD sensor. The first type of image noise consists of color dots due to the individual pixels having slightly different readouts: this becomes particularly visible in areas where the color is homogenous (chrome noise, Fig. 5.13). A second type of image noise consists of monochromatic variations of light intensity between neighboring pixels (luminance noise, Fig. 5.14).

The main systematic cause of error affecting the images relates to the fact that each pixel of a CCD sensor has slightly different gain. Therefore, if the same composition of light wavelength and intensity reaches different pixels, each will systematically read a slightly different value.

Random noise in the images is caused by a number of factors. First, the detection of photons by the CCD sensor is a statistical process. The intensity (number of photons detected) will vary slightly between images taken over subsequent periods of identical

Fig. 5.13 Chrome noise in an image acquired by a digital camera. Pixels with slightly different color appear in regions where the original color is indeed uniform

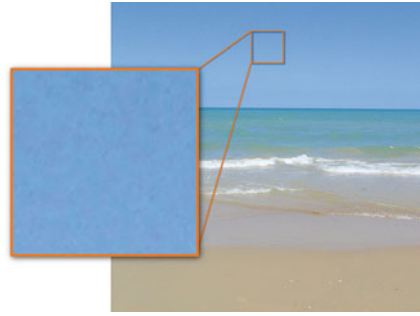
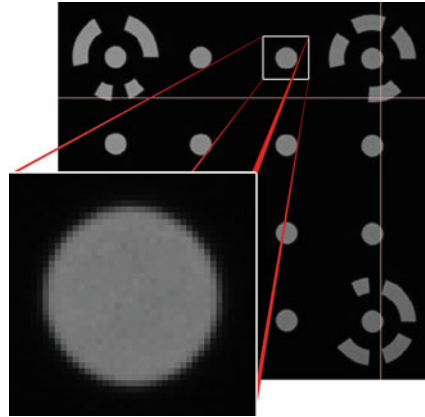


Fig. 5.14 Luminance noise in an image acquired by a digital camera. The detail illustrates the effect of monochromatic noise in the grayscale variations between adjacent pixels. Courtesy of Giacomo Lionello (Rizzoli Orthopaedic Institute, Bologna, Italy)



duration (the distribution of intensity between theoretically identical images follows the Poisson distribution). Secondly, additional electrons are generated within the CCD sensor not by the absorption of photons (i.e. the image signal), but by physical processes within the CCD itself, which are related to thermal vibration. The number of electrons generated in a given time depends on the operating temperature of the CCD. This type of noise is known as thermal noise (or dark noise), and can be reduced by cooling down the CCD sensor. An additional source of error is the so-called excess noise, having a pink wavelength (i.e. in the low frequencies in the visible spectrum). The excess noise has some memory effect, which makes it deviate from a zero average. While the previous sources of noise can be somewhat attenuated by filtering, excess noise is much harder to eliminate.

Finally, additional systematic and random errors appear in the subsequent signal processing, during the amplification and analog-to-digital conversion of the signal from each pixel. The consequences of this combination of errors affecting the acquired images propagate to the computation of the displacement fields. This step is generally quite robust, and shows relatively small noise. Conversely, when the displacement field is derived to calculate the strain distribution, errors tend to increase: if no filtering is adopted, even a uniformly strained surface will show an apparent strain with significant local variations noise, Fig. 5.15.

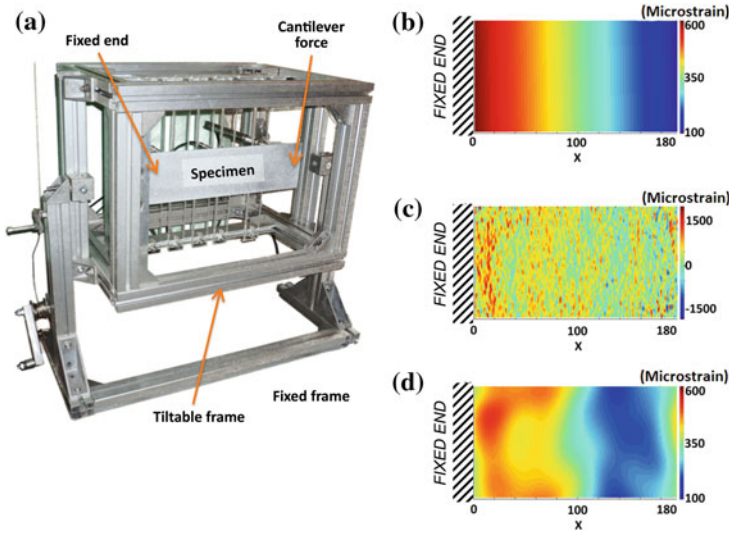


Fig. 5.15 Test arrangement to produce known displacements and strains: **a** a flat aluminum bar was subjected to a cantilever load, inducing a maximum strain of $f 750 \mu\epsilon$ at the fixed end; on the opposite side to the DIC camera, strain gages and LVDTs were placed to measure the actual displacements and strain. Theoretical strain distribution derived from the beam theory and the readouts from the LVDTs **(b)**. DIC-computed strain distribution with no filtering **(c)**, and after Gaussian low-pass filtering **(d)**

5.8.4 To Filter, or Not to Filter?

The main limitation of Digital Image Correlation (DIC) is the remarkable noise affecting the DIC-predicted strain distributions. To reduce noise, filters are often applied. However, this is also associated with loss of information (smoothing of the strain gradients).

Manufacturers of commercial DIC systems seldom provide guidelines for optimal filtering of DIC strain distributions other than *go on filtering until it looks like you expect it to be*. This is clearly the opposite of a scientific process, as it can virtually prevent discovering anything that is not expected (e.g. a high strain gradient). Therefore, it is necessary to systematically explore different filtering strategies to reduce noise and minimize the loss of information in the DIC-predicted strain distributions [12].

To provide an evidence-based indication of the most suitable filtering strategy, a preliminary analysis to measure the features of the noise to be removed must be carried out. This enables identifying the features of dark noise affecting the images acquired by the CCD cameras.

A first strategy one could consider consists of filtering the input images with median adaptive low-pass filters, or notch filters to eliminate the dark noise. However, both strategies actually produce an increase (rather than a reduction) of noise in the computed strain distributions.

A frequently used strategy is to apply a Gaussian low-pass filter to the strain distributions. When the optimal cutoff frequency is selected, the noise is remarkably reduced (by 70%) without excessive loss of information. At the same time, when non-optimal cutoff frequencies used, the residual noise and/or loss of information seriously compromised the results, Fig. 5.16.

A more effective strategy relies on filtering the DIC-computed displacement field: this low-pass filtering strategy incorporates a continuum assumption and is somewhat less arbitrary than filtering the strain distribution just because it does not look good.

Finally, image combination techniques can be applied both to the input images, and to the strain distributions. This strategy is extremely time-consuming but not very effective (noise is reduced by less than 10%).

In conclusion, it has been shown that the only truly effective filtering strategy is the one applied to the strain distribution [12]. An optimization of the cutoff frequency is possible (and mandatory) to achieve reliable results.

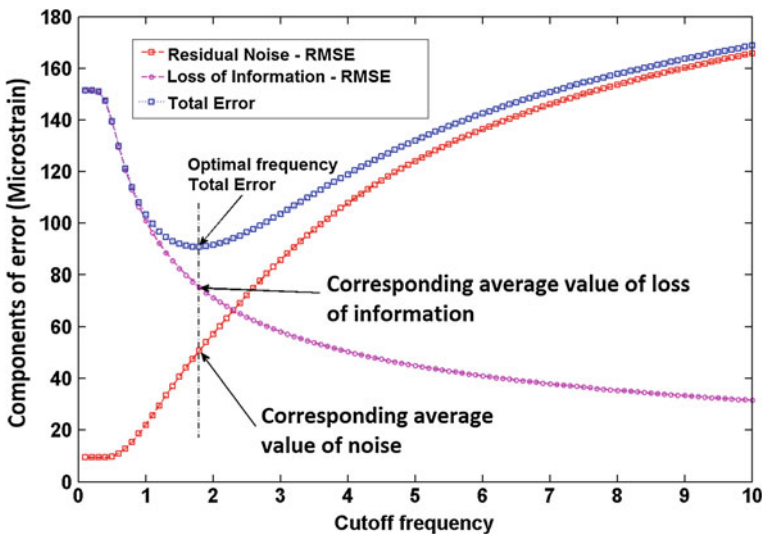


Fig. 5.16 Variation of the two terms composing the total error after filtering (residual noise, loss of information) as a function of the cutoff frequency. The longitudinal strains in a cantilever plate, Fig. 5.15 were processed, when a deflection of was applied at the free end (corresponding to a maximum strain of $750 \mu\epsilon$ at the fixed end). Courtesy of Jacopo Baldoni

5.9 Case Study: Optimization of the Hardware and Software Parameters

Before using the DIC to draw any conclusions, a focused optimization is necessary to obtain the best results (minimum systematic and random error). The starting point was an average systematic error of $100 \mu\epsilon$, and a noise of $500 \mu\epsilon$. At this stage, DIC is neither sufficiently accurate nor precise to measure in rigid specimens where strain does not exceed $2000 \mu\epsilon$. Such errors are much worse than the performance of strain gages.

The first step was optimizing the preparation of the specimen surface. The aim was to define the optimal airbrush settings to produce the optimal pattern for five different measurement windows chosen for biomechanical specimens. Different patterns were produced, to identify the optimal settings of the airbrush to generate speckles of the desired size, with minimal scatter [8].

The following software parameters were then examined: (i) size of the subsets of pixels (facet size), (ii) overlap between facets, (iii) filtering (contour smoothing). Following the design of the experiment (DOE) [13], a factorial design was used to establish a link between the software parameters and the errors. An unloaded flat specimen, Fig. 5.15 with a dedicated pattern for the specific measurement window was used to evaluate the effects produced by the alteration of software parameters on the strain measurement (as the specimen was not subjected to any real strain, any readout different from zero is an indicator of measurement error).

Subsequently, the strain distribution in a specimen with a simplified geometry (cantilever load applied to the same aluminum flat specimen) was measured using DIC, while two strain gages were bonded on the opposite face, as a reference.

In the end, the hardware parameters were explored in order to minimize noise: (i) the camera gain, (ii) the shutter time. Using the results of the previous tests, it was possible to investigate the effect produced by the alteration of the hardware parameters. The same unloaded flat specimen was used in these tests. In the first test, different gain levels and a fixed exposure were used. The aim was evaluate the effects produced by the alteration of gain level on the errors affecting the measured strain. In the second test, the best gain level was fixed, and the optimal exposure was evaluated to obtain the minimum error and noise.

Remarkable improvements in terms of accuracy and precision were achieved, Fig. 5.17. Starting from an average systematic error of $100 \mu\epsilon$, and a noise of $500 \mu\epsilon$, using the knowledge developed during these tests, the error decreased to $10 \mu\epsilon$ and the noise to $100 \mu\epsilon$.

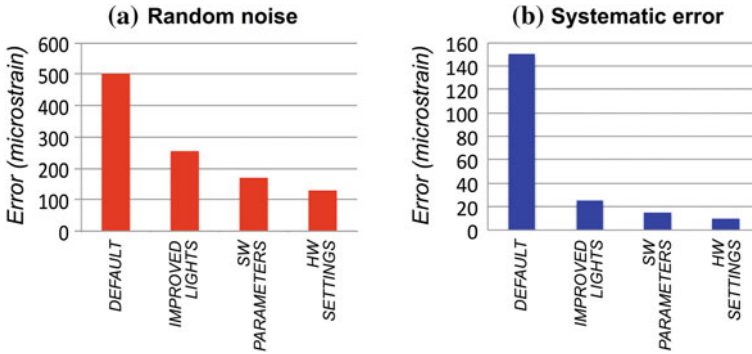


Fig. 5.17 A flat specimen was mounted on a tiltable frame to test the DIC system, Fig. 5.15. Its surface was analyzed in both a-zero strain condition, and when a cantilever load was applied. Plots **a** and **b** show how the random noise and the systematic error were dramatically high if no special care was taken, and how errors can be reduced if the hardware and software settings are optimized following the principles of DOE

5.10 Case Study: Strain Distribution in the Human Tibia

The strain distribution in the human tibia is particularly puzzling as it shows some peculiar optimization [14]. In this study the displacements and strains in the human tibia were investigated when a cantilever load was applied, Fig. 5.18.

5.11 Digital Volume Correlation

Recently, the power and resolution of diagnostic imaging has improved, so that the distribution of material properties inside a structure can be accurately measured. In this case, the three-dimensional images are constituted by voxels. Contemporary

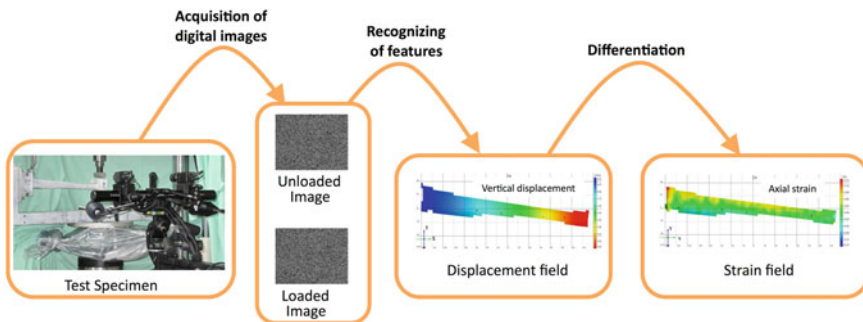


Fig. 5.18 Experimental setup: a cantilever load was applied to a human tibia prepared with a speckle pattern while DIC recorded stereoscopic images. From the speckle pattern, the displacement and strain fields were obtained. (Copyright of the LHDL consortium; reproduced with permission)

techniques include: computed tomography (CT, with a typical voxel of 0.1–1 mm), high-resolution computed tomography (micro-CT, with a resolution down to some micrometers), magnetic resonance (MRI, with a resolution of some millimeters) and high-resolution-MRI (with a resolution of some tens of micrometers), and positron emission tomography (PET).

Such imaging techniques, in combination with in situ mechanical testing, has enabled the measurement of 3D full-field displacement and strain fields throughout the interior of a material by comparison of an underformed and a deformed state. This technique is called Digital Volume Correlation (DVC) and its operating principle is an extension of the DIC.

One of the very first implementations of DVC was proposed by Bay et al. in 1999 to investigate the strain distribution throughout a complex biological structure, such as the trabecular bone [15]. Since then, the technique has gained an increasing attention in the fields of biomechanics and biomaterials; but also in experimental mechanics and material research related to various engineering applications [16]. DVC can exploit more structural information than DIC, and can therefore be coupled with finite element solvers for the calculation of the internal strain field, while applying a continuum assumption [17].

In order to be investigated by means of DVC a structure needs to have suitable materials properties for the selected imaging technique (e.g. radiopacity if micro-CT is used). At the same time, the material must possess an internal inhomogeneous structure that can be viewed by such a technique, and exploited as a natural 3D speckle pattern, with similar limitations and requirements as the pattern for surface inspection by means of DIC (dimensions, contrast, etc.). For instance, biological tissues such as bone offer a suitable internal pattern thanks to the porous trabecular structure, Fig. 5.19. Similarly, the fibres/particles of composite materials can serve as a tracking pattern. Other materials suitable for DVC are cast iron (thanks to the internal graphite particles), and inhomogeneous rocks and soils.

A major limitation is the very long time required to acquire 3D images of suitable quality: a micro-CT scan with adequate resolution and quality can require between 20 min and some hours. Therefore, images are acquired under stationary conditions, after applying a known load or strain to the specimen. It is clear that if the material under investigation is viscoelastic, deformational and failure will be affected by such a slow procedure. Although DVC is a very promising technique, it should always be used with extreme caution. In fact, the accuracy and precision of the predicted displacements and strains cannot be taken for granted [17–19]. An intrinsic difficulty in validating DVC is that currently no other technique exists that allows measuring strains inside a loaded structure. Therefore, no golden standard is available to measure the accuracy and precision of DVC. While the calculated displacements are in general very accurate (with a precision and accuracy better than the voxel size), the computed strains can be affected by large errors (up to 50,000 $\mu\epsilon$; even with the best settings, errors exceed several hundreds $\mu\epsilon$) [19].

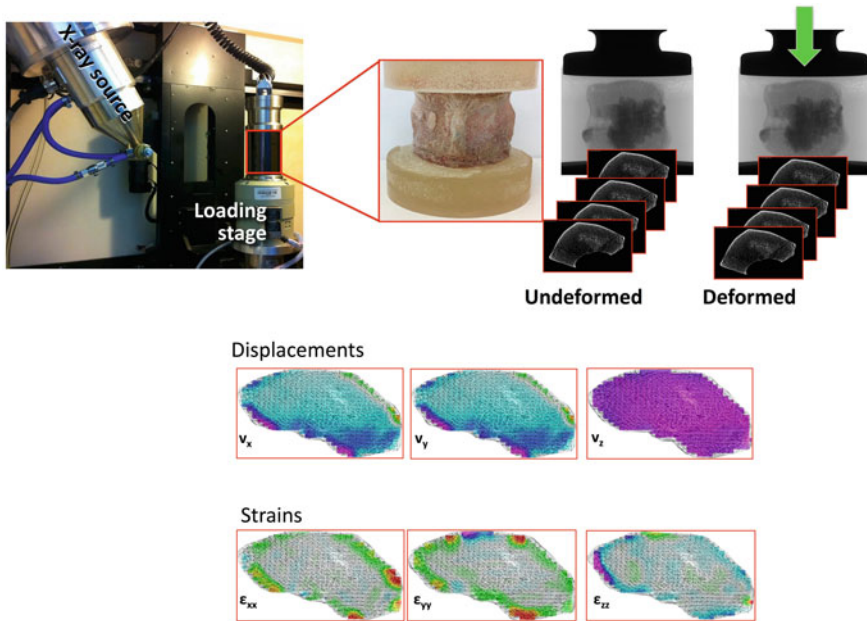


Fig. 5.19 Example of application of DVC to the mechanical characterization of a vertebra. The specimen was loaded in a high-resolution CT. Correlation of the undeformed and deformed volumetric images enabled computing the distribution of displacements and strain. Courtesy of Dr Gianluca Tozzi (University of Portsmouth, UK)

References

1. Peters W, Ranson W (1982) Digital imaging techniques in experimental stress analysis. *Opt Eng* 21(3):427–431
2. Luo PF, Chao YJ, Sutton MA, Peters WH (1993) Accurate measurement of three dimensional deformations in deformable and rigid bodies using computer vision. *Exp Mech* 32(2):123–132
3. Schreier HW, Braasch JR, Sutton MA (2000) Systematic errors in digital image correlation caused by intensity interpolation. *Opt Eng* 39:2915–2921
4. Dally JW, Riley WF (2005) *Experimental stress analysis*, 4th edn. College House Enterprises Llc, Knoxville
5. Sutton MA, Orteu JJ, Schreier H (2009) *Image correlation for shape motion and deformation measurements: basic concepts theory and applications*. Springer Science and Business Media, New York
6. Ajovalasit A (2009) *Analisi sperimentale delle tensioni con la fotomeccanica: Fotoelasticita, moire, olografia speckle, correlazione immagini*, 2nd edn. Aracne Editrice, Rome
7. Barranger Y, Doumalin P, Dupre J-C, Germaneau A (2010) Digital image correlation accuracy: influence of kind of speckle and recording setup. *Eur Phys J Web Conf* 6:31002
8. Lionello G, Cristofolini L (2014) A practical approach to optimizing the preparation of speckle patterns for digital-image correlation. *Meas Sci Technol* 25(10):107001
9. Reu P (2012) Hidden components of DIC: calibration and shape function—part 1. *Exp Tech* 36(2):3–5
10. Zhou P, Goodson KE (2001) Subpixel displacement and deformation gradient measurement using digital image/speckle correlation. *Opt Eng* 40(8):1613–1620

11. Clocksin WF, Quinta da Fonseca J (2002) Image processing issues in digital strain mapping. *Proc SPIE* 4790:384–395
12. Baldoni J (2013) Riduzione del rumore nelle misure di deformazione tramite correlazione digitale di immagini. Master Thesis Presented at the School of Engineering and Architecture, University of Bologna. <http://amslaurea.unibo.it/5559/>
13. Montgomery DC (2005) Design and analysis of experiments, 6th edn. Wiley, New York
14. Cristofolini L, Angeli E, Juszczuk JM, Juszczuk MM (2013) Shape and function of the diaphysis of the human tibia. *J Biomech* 46(11):1882–1892
15. Bay BK (2008) Methods and applications of digital volume correlation. *J Strain Anal Eng Des* 43(8):745–760
16. Roberts BC, Perilli E, Reynolds KJ (2014) Application of the digital volume correlation technique for the measurement of displacement and strain fields in bone: a literature review. *J Biomech* 47(5):923–934
17. Dall’Ara E, Barber D, Viceconti M (2014) About the inevitable compromise between spatial resolution and accuracy of strain measurement for bone tissue: A 3D zero-strain study. *J Biomech* 47(12):2956–2963
18. Liu L, Morgan EF (2007) Accuracy and precision of digital volume correlation in quantifying displacements and strains in trabecular bone. *J Biomech* 40(15):3516–3520
19. Palanca M, Tozzi G, Cristofolini L, Viceconti M, Dall’Ara E (2015) 3D local measurements of bone strain and displacement: comparison of three digital volume correlation approaches. *J Biomech Eng (ASME)* (in press)

Part II

Developing Phenomenological Theories and Problem Solving

In this second part of the book the authors give an overview of the main fields of Structural Mechanics to show the extent and the effectiveness of experimental stress analysis for the understanding of phenomena and assessment of procedures.

In order to show the interaction between theoretical models and experimental data, each chapter offers a sufficient introduction to the main concepts and variables, without overwhelming the reader with an excess of specialization. For more detail, the reader is referred to more specialized publications, some of which can be found in the References.

Chapter 6

Static Stress Models

Abstract Experimental methods to develop static stress analysis models might appear as a superfluous option since numerical analysis has reached so high a level of refinement and accuracy to cover practically every kind of structural mechanics requirements. The field of Statics was the first to take advantage of numerical methods. Nevertheless, experiments continue to be fundamental for building simulation models. A few case studies with classic applications of experimental methods to pressure vessels—the old history of experimental mechanics—are followed by elementary cases of identification of unknown variables, with the purpose of showing the potential of the inverse approach. Models for the simulation of constitutive materials laws are shown for stress states under and beyond the elastic limit. The reciprocal influence of the stress state and the behavior of the material is discussed. A brief note on the classic elementary models of the physical theory of fracture is finally given.

6.1 The Illustrative Advantage of a Full Field Analysis

The experimental analysis, if from one hand confirms and verifies theoretical results, on the other hand shows the limits on applicability of the theory, e.g. of the Saint Venant simplified theory about the concept of *slender bodies* or of the contact problems where the load is transferred between two bodies through small surface areas. For a general reference, the reader can find in following books the fundamentals of the elastic and inelastic theory [1–5]. As first reference [6].

6.1.1 Force at a Point of a Straight Boundary

This classical Boussinesq theory covers a frequent design praxis, when elements are conceived to transmit forces from one another, with localized actions on small areas.

A concentrated load acting perpendicularly to an half plane, Fig. 6.1 is the general reference scenario of every similar structural solution. The relationship between

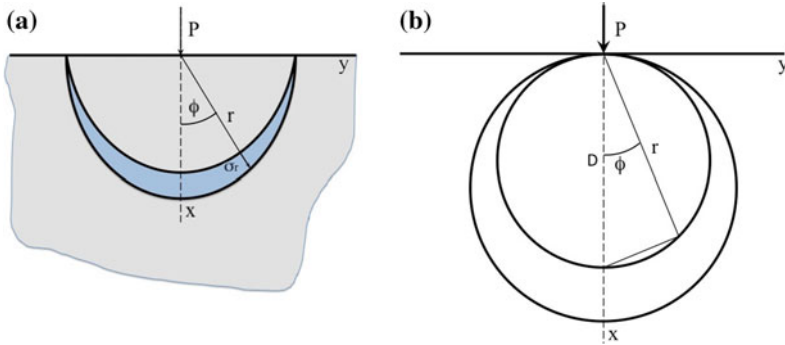


Fig. 6.1 Concentrated load perpendicular to a half-plane: **a** radial stress distribution, **b** contour curves of the radial stress

stresses and load is:

$$\sigma_r = -\frac{2P \cos \phi}{\pi d} \frac{1}{r} \quad \sigma_\phi = 0 \quad \tau_{r\phi} = 0 \quad (6.1)$$

The radial stress is a principal stress at any point while the other is zero. From Eq. 6.1 is obtained:

$$\frac{\cos \phi}{r} = \text{const.} \quad (6.2)$$

The loci of constant shear stress are tangent at the contact point and have a circular shape with centers on the normal x axis. The constant of the Equation, determined by the condition $\phi = 0$, is equal to $1/D$.

Exercise 6.1 (*Stress distribution due to a force on boundary of a disk*) The exercise deals with the possibility to assume the stress distribution of a semi-infinite body for the approximated determination of the stress state in a circular disk with a concentrated load at a point of its boundary. It utilizes the photoelastic method: by this analysis, Chap. 3, it is possible to verify that isochromatics (loci of constant shear stress) are loci of constant radial stress.

(a) As first step, given the photoelastic fringes of Fig. 6.2a for a disk, the unknown concentrated load is approximately identified, assuming the Boussinesq’s formula of a semi-infinite body.

Data:

- Disc diameter $2a = 60$ mm.
- Thickness $d = 10$ mm.
- Constant of fringes of *Araldite B* $f_\sigma = 10.2$ [N/ord.mm].

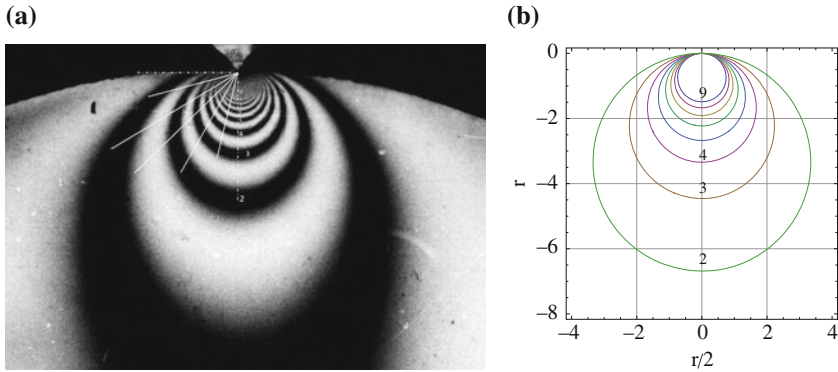


Fig. 6.2 a Patterns of a photoelastic analysis of the stress distribution in a concentrated contact and b theoretical model of the approximated model of Eq.6.3

Since the dimensions of variables are: [Newton/ord.mm] for f_σ , mm for r , P will be identified in *Newton*. Equation is:

$$\sigma_r = -\frac{2P \cos \phi}{\pi d} \frac{1}{r} \quad \sigma_\phi = 0 \quad \tau_{r\phi} = 0 \quad (6.3)$$

Equation 6.3 can be rewritten as¹:

$$N_{ord} = -\frac{2P \cos \phi}{\pi} \frac{1}{r} \cdot \frac{1}{f_\sigma}$$

Figure 6.3 shows (a) the experimental data (35 points) in cylindrical coordinates r and ϕ , fitted by the Eq.6.3 through an interpolation algorithm and (b) the model applied for fitting only to the experimental data derived on the symmetry axis for $\phi = 0$.

The identified value of the load is:

$$P = -178 \text{ N}$$

The approximation of the theoretical model to interpolate the experimental data shows an acceptable use of the theory of infinite half-plane, almost for points not too far from the contact area ($r \leq 6$ mm). At larger distance, stresses are influenced by the real boundary conditions of the disk that cannot be assimilated to an infinite plane.

¹ The general equation of photoelasticity, Chap. 3, is:

$$\frac{\sigma_1 - \sigma_2}{f_\sigma} = \frac{N_{ord}}{d}$$

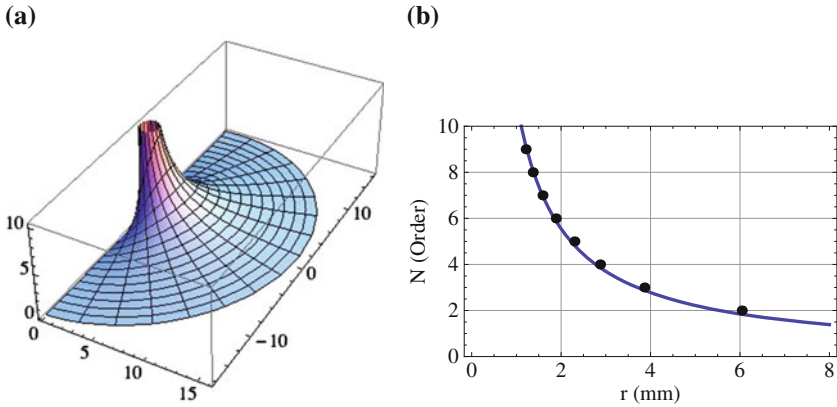


Fig. 6.3 Theoretical reconstruction by Eq. 6.3 in (a) and the same theoretical model fitted to the experimental data for $\phi = 0$ in (b)

6.1.2 Hertzian Contact

The Hertz theory of contacts is a nice application of the theory of elasticity [6].

In the case of a contact between two cylinders of radii R_1 and R_2 with parallel axes, the contact surface is rectangular with one side equal to the cylinders thickness d and the other side equal to a finite value $2b$, caused by the local deformation. If R_2 tends to infinity, we have the case of Fig. 6.4 of the Hertzian contact between a cylinder and an half-plane.

The contact stress distribution q along $2b$ is semi-elliptically distributed with a maximum central value equal to q_0 . $P' = P/d$ is the load per unit thickness along the x axis, perpendicular to the figure plane. The y and z axes are placed along horizontal and vertical directions. The link between P' and q_0 is given by the equation:

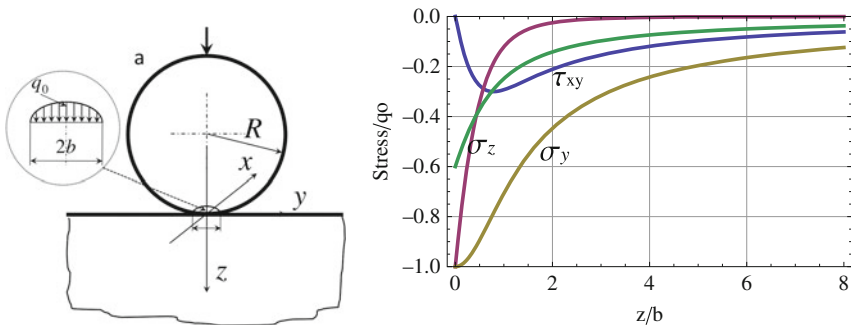


Fig. 6.4 Diagram of stress components along the depth for the hertzian contact between a cylinder and an half-plane for $\nu = 0.3$

$$q_0 = \frac{2P'}{\pi b} \quad (6.4)$$

The analysis of local deformation gives the following expression for the half-width of the contact segment b :

$$b = \sqrt{\frac{4P'(k_1 + k_2)R_1R_2}{R_1 + R_2}} \quad (6.5)$$

where R_1 and R_2 are the radii of the cylinders and k_1 and k_2 are elastic constants, defined by:

$$k_1 = \frac{1 - \nu_1^2}{\pi E_1} \quad k_2 = \frac{1 - \nu_2^2}{\pi E_2} \quad (6.6)$$

If both cylinders are of the same material with $\nu = 0.3$, Eq. 6.5 becomes:

$$b = 1.52\sqrt{\frac{P'R_1R_2}{E(R_1 + R_2)}} \quad (6.7)$$

If cylinders are also geometrically identical ($R_1 = R_2 = R$), then:

$$b = 1.08\sqrt{\frac{P'R}{E}}$$

For contact between a cylinder and a plane Eq. 6.7 gives:

$$b = 1.52\sqrt{\frac{P'R}{E}}$$

The value of the maximum pressure q_0 is obtained by substituting Eq. 6.5 in Eq. 6.4:

$$q_0 = \sqrt{\frac{P'(R_1 + R_2)}{\pi^2(k_1 + k_2)R_1R_2}} \quad (6.8)$$

If the materials of both cylinders are identical and $\nu = 0.3$, the maximum pressure is:

$$q_0 = 0.418\sqrt{\frac{P'E(R_1 + R_2)}{R_1R_2}} \quad (6.9)$$

while for the contact between a cylinder and a plane is:

$$q_0 = 0.418\sqrt{\frac{P'E}{R}} \quad (6.10)$$

From the expressions of q_0 and b , the stresses in any point are determined [6, p.404].

Figure 6.4 shows the stress distribution for $\nu = 0.3$. The maximum shear stress is equal to $0.304 q_0$ and is not localized on the surface, but at a certain depth from the area of contact along the z axis ($z_1 = 0.78 b$). This result has implication in modeling the phenomenon of surface fatigue with repeated loads.

In case of contact between a cylinder and a plane with different elastic materials the expressions of b and q_0 of Eqs. 6.5 and 6.8 become:

$$b = \sqrt{4P'(k_1 + k_2)R} \quad (6.11)$$

$$q_0 = \sqrt{\frac{P'}{\pi^2(k_1 + k_2)R}} \quad (6.12)$$

This formula must be considered for the analysis of a photoelastic experiment with disc and plane of different materials.

The theory can be completed with the calculation of the distributions of σ_x , σ_y and σ_z , Fig. 6.4b. All the previous normal stresses are compressive and principal for symmetry reasons:

$$\begin{cases} \sigma_x = -2\nu q_0 \left(\sqrt{1 + z^2/b^2} - \frac{z}{b} \right) \\ \sigma_y = -q_0 \left[\left(2 - \frac{1}{1 + z^2/b^2} \right) \sqrt{1 + z^2/b^2} - 2\frac{z}{b} \right] \\ \sigma_z = \frac{-q_0}{\sqrt{1 + z^2/b^2}} \end{cases} \quad (6.13)$$

and τ_{max} becomes:

$$\tau_{max} = \frac{\sigma_z - \sigma_y}{2} = \frac{q_0}{2} \left[\left(2 - \frac{1}{1 + z^2/b^2} \right) \sqrt{1 + z^2/b^2} - 2\frac{z}{b} - \frac{1}{\sqrt{1 + z^2/b^2}} \right] \quad (6.14)$$

The maximum shear stress (isochromatic with maximum order) is under the surface contact at a depth that is function of the load and radius. For $\nu = 0.3$ this depth is $z_1 = 0.78b$ and the maximum shear stress is $0.304 q_0$.

Exercise 6.2 (*Contact stress between a cylindrical disk and a plane surface*) Identify the maximum value of the pressure q_0 in a hertzian contact between the disc in plastic material and the plane surface in steel, being given the photoelastic fringes distributions in a transparent plastic cylindrical disk.

(a) Data:

- Disc diameter $2a = 60$ mm
- Elastic modulus of disc material $E_1 = 3,100$ MPa

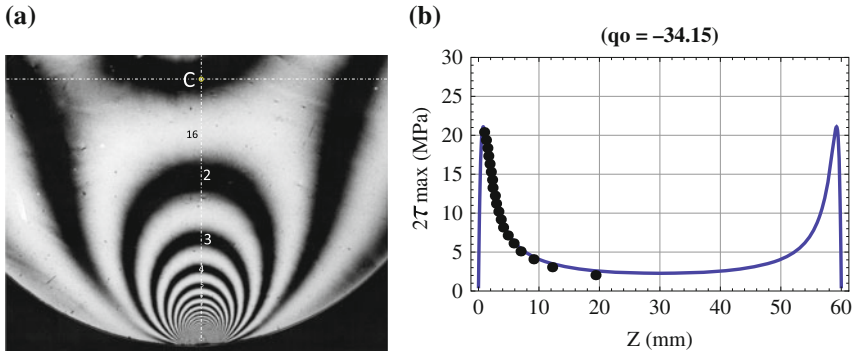


Fig. 6.5 Photoelastic fringes (a) in the contact between the photoelastic disc and the plane steel surface, fitted by theoretical model (b) of Eq. 6.15

- Elastic modulus of half-plane material $E_2 = 210,000 \text{ MPa}$
- Poisson’s coefficient for both materials $\nu = 0.3$
- Photoelastic constant for *Araldite B* $f_\sigma = 10.2 \text{ [N/ord.mm]}$
- Photoelastic fringes given in Fig. 6.5.

(b) The theoretical model that relates the fringe orders to the maximum pressure in the contact area is given by Eq. 6.14 for a contact between a disc and an half plane. Being the orders of fringes proportional to $(\sigma_z - \sigma_y)$, it is possible to identify the load in the disk that is optically sensitive.

For this task the Eq. 6.14 can not be directly applied because it was developed for an half plane. Nevertheless, for the high localized phenomenon of contact stress it can be approximately utilized also for a disk but the model valid for an half-plane must be modified for taking into account the double loading application on the disk in a radial direction. For this reason, instead of Eq. 6.14 the new Eq. 6.15 was developed that sums two opposite loading contributions at a distance $2a$.

$$\begin{aligned} \sigma_z - \sigma_y = q_0 \left[\left(2 - \frac{1}{1 + z^2/b^2} \right) \sqrt{1 + z^2/b^2} - 2\frac{z}{b} - \frac{1}{\sqrt{1 + z^2/b^2}} \right. \\ + \left(2 - \frac{1}{1 + (-z + 2a)^2/b^2} \right) \sqrt{1 + (-z + 2a)^2/b^2} \\ \left. - 2\frac{(-z + 2a)}{b} - \frac{1}{\sqrt{1 + (-z + 2a)^2/b^2}} \right] \end{aligned} \tag{6.15}$$

(c) By an interpolation algorithm the value of q_0 is identified (-34.15 N/mm), fitting this model to experimental data (Sign minus means a compressive state).

(d) P' is a function of q_0 and is given by Eq. 6.8: $P' = 32.77 \text{ N/mm}$ and $b = 0.61 \text{ mm}$ (by Eq. 6.4).

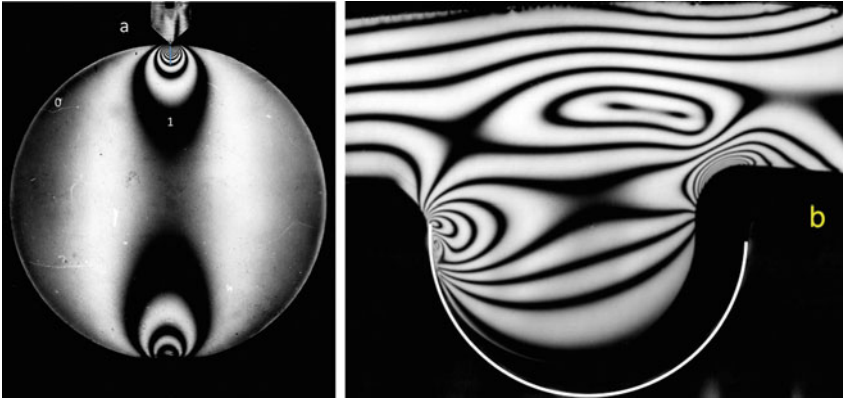


Fig. 6.6 Photoelastic fringes of contact stresses in a disc due to **a** a concentrated load (*upper side*) and a half-plane (*lower side*), and **b** in a toothed belt acting in a pulley in a transmission device

All the examples confirm the validity of Saint Venant principle: stresses in the upper and in the lower part of the disc at a certain distance from the applied loads, Fig. 6.6 have the same distribution.

In spite of its clarification potential, the classical photoelasticity is overruled by numerical methods, especially for solving forward problems. It maintains value for inverse problems. The inherent limit due to a different material of the model and of the real object (which makes it possible only an extended similarity in the case of contact problems, see [7]), is overcome in special applications where the plastic materials are utilized for structural solutions, as composites or elastomeric elements for o-ring seals or toothed belts. Figure 6.6, shows the contact singularities in a toothed belt inside a metal pulley, overlapped with the general loading effect.

6.2 Design by Experiments of Pressure Vessels

If the shape of the body and the loads distribution are complex with three-dimensional stress states (e.g. pressure vessels), stress concentration factors, typical of slender bodies with elementary load conditions can not be extended and utilized and there is no reference to classical handbooks as [8] for estimating the peak stresses at critical points. In this case stress concentration in geometrical discontinuities can only be evaluated numerically or experimentally.

Stress state concentration was one of the main task of experimental stress analysis when numerical methods were not so common as today are. This original task of

the experimental approach (when to verify the approximate numerical estimations was recommended), is partially overruled, even if the designer choices for structural optimization often take advantage of analyses on physical models or on prototypes themselves.

6.2.1 Case Study of Steam Generator Reactor Head with Four Openings on the Spherical Dome

Limiting the analysis only to some cases of special interest or complexity, it is possible to verify some advantages of experimental methods. A photoelastic analysis was carried out on a generator head. In Chap. 3 the casting form of the model is shown; in Fig. 6.7 the model and fringes in a symmetry plane are shown. Figure 6.8 shows the results of the analysis: the diagram of the normal stresses along the boundary (σ_T and σ_N perpendicular to the symmetry plane), both normalized respect to the membrane stress in the spherical part ($\sigma_0 = pD/2h$ where D is the mean sphere diameter and h the thickness).

6.2.2 Case Study of a Reactor Pressure Vessel with Four Nozzles in the Cylindrical Part

A reactor pressure vessel is typically constituted by a cylindrical shell with radial nozzles and a spherical head fastened to the main body by means of a flange. Figure 6.9 shows a sketch of the reactor container with the symmetry critical plane along the longitudinal symmetry plane of a nozzle, where the maximum circumferential stress is located.

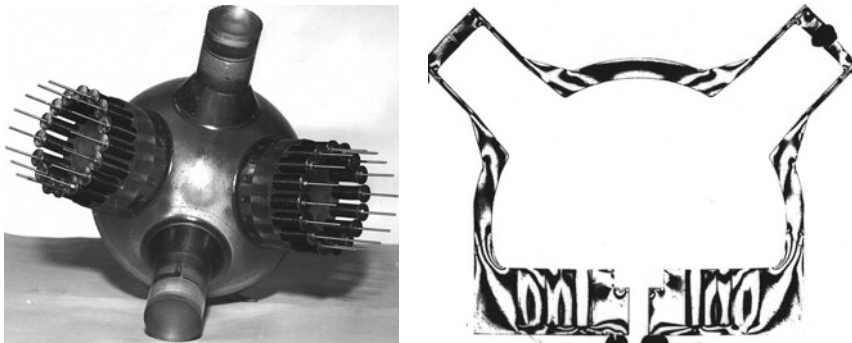


Fig. 6.7 Monolithic model of epoxy resin and interferometric fringes in one of the symmetry planes of a pressure vessel with four nozzles; the corresponding casting mold in Chap. 3

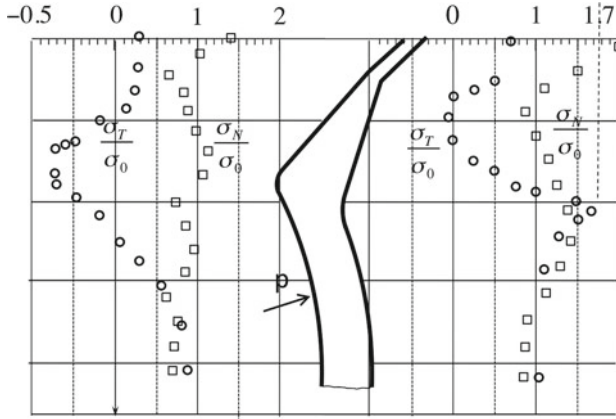


Fig. 6.8 Perpendicular and tangential normal stresses on the symmetry plane

Loadings are given by internal pressure, forces and moments applied to the nozzles and tightening load on the flange. Figure 6.10 shows a comparison of circumferential stress distributions (i.e. perpendicular to the figure plane) due to an internal pressure, on the inner surface of the vessel, for four different designs, [9] for details and bibliography. It is worth observing that the normalized values of the circumferential stresses σ_N/σ_0 in the area in the inner side, between the cylindrical vessel and the nozzle is influenced by few geometric dimensionless factors, Table 6.1. In other words a local optimization of the peak stress is possible by small adjustment of few factors.

This is a confirmation of an analysis that identifies the few contributions that account for most results. It gives a measure of significance to factors that may not appear always significant at first, such as t'/T' .

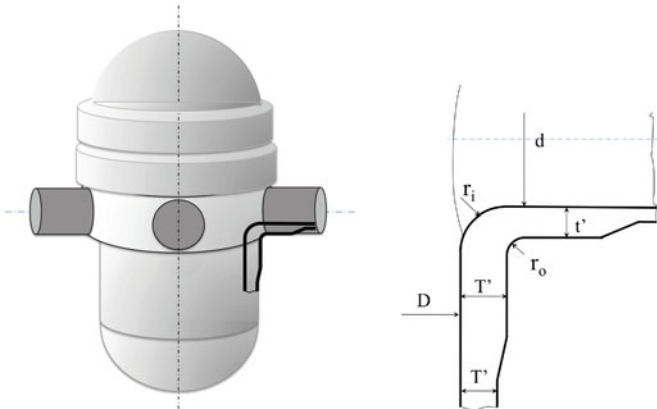


Fig. 6.9 Pressure vessel and nozzle sizes

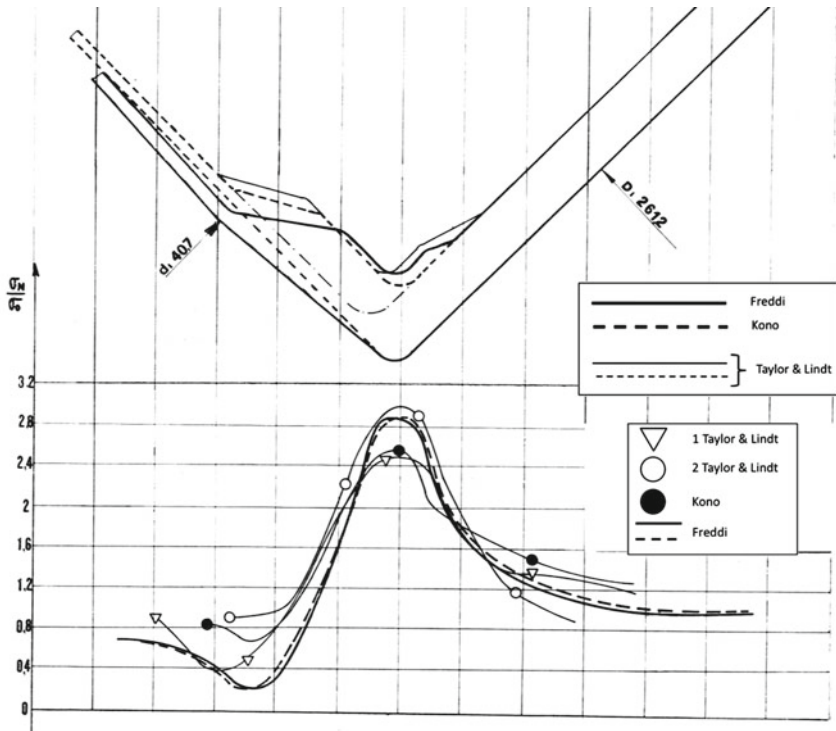


Fig. 6.10 Circumferential stress distribution in longitudinal plane for different but similar geometries; data in Table 6.1

6.3 Stress Concentration Factors

The normalization of the stress diagrams introduced in the previous subsections in the pressure vessels studies (stress at a point is expressed as a dimensionless ratio) is only a useful representation of the stress diagrams in complex three-dimensional systems, but does not allow full generalization of these values.

Vice-versa, when a component can be treated by the elementary Saint Venant theory, the concept of the concentration factors is more general and can be collected

Table 6.1 Comparison of stress ratio σ_N/σ_0 of similar nozzles as function of dimensionless variables (note that $s/S = dT'/Dt'$)

Nozzles	D/T	t'/T'	d/D	s/S	r_i/T'	r_o/T'	$K(\sigma_N/\sigma_0)$	Note
Experim.	17.4	1.00	0.226	0.226	0.27	0.50	2.8	Phot. [9]
Numerical	17.4	1.00	0.226	0.226	0.27	0.50	2.8	Fin.el [10]
Experim.	14.7	0.76	0.193	0.254	0.43	0.43	2.52 (2.22)	Phot. [11]
Experim.	15.6	1.00	0.285	0.285	0.33	0.44	2.84 (2.55)	Phot. [9]

in specialized handbooks [8, 12], and utilized for a variety of similar geometries. The stress concentration factors are utilized in *fatigue* studies, since stress concentrations due to geometrical discontinuities are the main source of reduction of components life when the stresses vary over time, see Chaps. 7 and 8. In case of static stress their role in augmenting the local stress and also locally (i.e. limited to small areas) exceeding of yield stress is not considered dangerous. In spite of this fact, the concentration factors are defined statically and determined by the theory of elasticity or estimated by numerical or experimental methods.

The geometrical discontinuities arise from the general outline of the body or from imperfections due to manufacturing technology (e.g. defects in the welding joints) and also from service conditions (such as certain concentrated forms of corrosion that creates surface cracks). However, as it will be seen in dealing with variable loads, it is wrong to think that strength reduction derives only by the stress increment in discontinuities.

As it will be argued in following chapters, the effect on fatigue resistance of notches, grooves, sections variations, cracks etc. is much more complex. To anticipate some concepts that will be treated in them, the fatigue resistance depends on a variety of factors not only connected to stress state such as:

- Material structure at the root of the notch (which may differ from the base material condition for local thermo-mechanical treatments).
- Actual values of stresses and strains (only in elastic field proportional to the load).
- Stresses and strain gradients.
- Local plastic flow.

Coming back to the determination of stress concentrations, in the case of slender structures (e.g. beams, thin shafts and thin plates) referring to the *saint-venantian* hypothesis, stress concentrations are interpreted by concentration factors K_t , for each type of internal elementary stress distribution (normal and shear force, bending and torsion moments). They are defined as:

ratios between maximum stress in the notch cross-section, according to a linear elastic theory and nominal stress in the same section, calculated by the elementary Saint Venant theory.

So defined the concentration factors are:

- Defined in elastic range.
- Independent of loads level.
- Independent of the absolute size.
- Dependent only on the kind of internal stress, as normal or shear forces, bending or torsion moments.

For this reason e.g. in the German technical literature the concentration factor is properly called *Form Factor* (with designation α_k).

6.3.1 Basic Theoretical Cases: Stress Concentration for Circular Holes in Thin Plates

A first typical case of geometrical discontinuity is offered by a circular hole in a thin plate of large dimensions, loaded in one direction by uniform tension at a certain distance from the hole, Fig. 6.11. For this case a classical and ingenious theoretical solution was due to Kirsch quoted in [6].

Before showing the final formula for this case, it is worth recalling the way to determine it. The compatibility equations in case of rectangular coordinates are:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) (\sigma_x + \sigma_y) = 0 \tag{6.16}$$

Introducing the *Hamilton* operator:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{6.17}$$

the previous equation can be written as:

$$\nabla^2 (\sigma_x + \sigma_y) = 0 \tag{6.18}$$

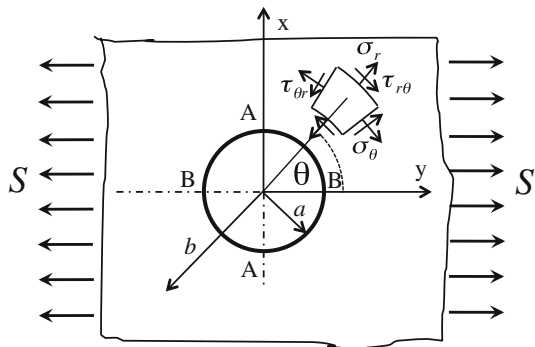
Airy showed that if a function $\phi = \phi(x, y)$ exists, such that the stress components can be written as:

$$\sigma_x = \frac{\partial^2 \phi}{\partial y^2} \quad \sigma_y = \frac{\partial^2 \phi}{\partial x^2} \quad \tau_{xy} = -\frac{\partial^2 \phi}{\partial x \partial y} \tag{6.19}$$

The equilibrium equations are automatically satisfied, because the following holds:

$$\frac{\partial \sigma_x}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} = 0 \quad \frac{\partial \sigma_y}{\partial y} + \frac{\partial \tau_{xy}}{\partial x} = 0 \tag{6.20}$$

Fig. 6.11 Circular hole in a loaded thin plate



The only condition for the function ϕ is continuity and three-times derivability respect to x and y . This function is purely mathematic and gives stress fields that not necessarily represent solutions for real problem, but can be used to describe many real problems, as the present one.

To determine the unique solution of the problem, Eq. 6.19 must also satisfy the previous compatibility equations Eq. 6.16:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \left(\frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial x^2} \right) = 0 \quad (6.21)$$

that, introducing the *Hamilton* differential operator, can be written as:

$$\nabla^4 \phi = \nabla^2 \nabla^2 \phi = \frac{\partial^4 \phi}{\partial x^4} + 2 \frac{\partial^4 \phi}{\partial^2 x \partial^2 y} + \frac{\partial^4 \phi}{\partial y^4} = 0 \quad (6.22)$$

This differential equation that gives a solution to the elastic problem when the boundary conditions are satisfy, is called biharmonic and functions ϕ that satisfy it are called biharmonic functions.

Boundary conditions can be specified in three different ways:

- Loads or stress given on the entire boundary.
- Displacements given on the entire boundary.
- Displacements given on a portion of the boundary and stresses given on the residual part of the boundary.

The first case is analyzed in the book of Alfirevic [13] where an important conclusion is emphasized:

Stress distribution does not depend on elastic material constants because they do not compare in Eq. 6.22 nor in the boundary conditions. For this reason experimental analysis can be carried out also on materials different from the material of the original structure. That is true for simply connected bodies or for multiply connected bodies, when the external loads on every closed boundary are in equilibrium, Fig. 6.12 (from [13]). In other case stresses depend on Poisson's coefficient.

Dealing with the problem of a circular hole, it seems natural to turn the previous equations into polar coordinates r, θ .

In rectangular coordinates x, y the components of stress vector in a direction θ is, for the analogy with Eq. 2.15 of Chap. 2²:

$$\sigma_{x'} = \frac{\sigma_x + \sigma_y}{2} + \frac{\sigma_x - \sigma_y}{2} \cos 2\theta + \tau_{xy} \sin 2\theta \quad (6.23)$$

² σ in place of ε , and τ_{xy} in place of $\gamma_{xy}/2$.

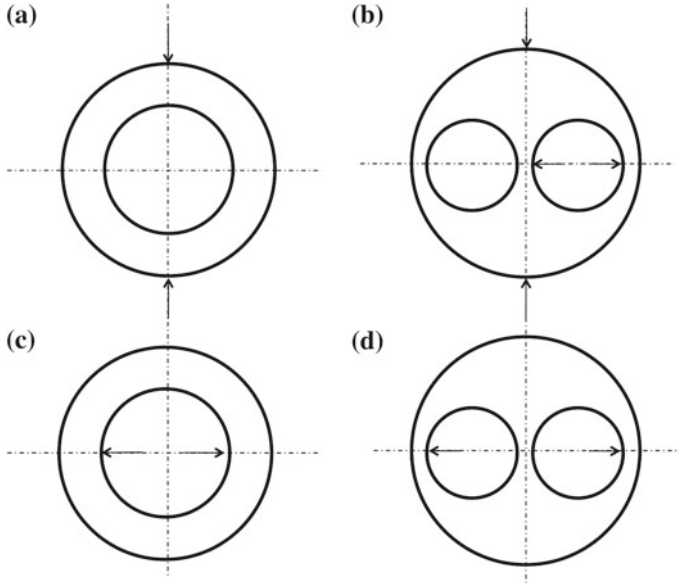


Fig. 6.12 Multiply connected bodies. For the cases **a**, **b** and **c** the stress distribution does not depend on elastic constants. Only for the case **d** stresses depend on Poisson’s coefficient (modified from [13])

Switching from rectangular coordinates to a polar reference system, the stress vector can be transformed in formal mathematical way and becomes:

$$\begin{cases} \sigma_r = \frac{\sigma_x + \sigma_y}{2} + \frac{\sigma_x - \sigma_y}{2} \cos 2\theta + \tau_{xy} \sin 2\theta \\ \sigma_\theta = \frac{\sigma_x + \sigma_y}{2} - \frac{\sigma_x - \sigma_y}{2} \cos 2\theta - \tau_{xy} \sin 2\theta \\ \tau_{r\theta} = -\frac{\sigma_x - \sigma_y}{2} \sin 2\theta + \tau_{xy} \cos 2\theta \end{cases} \quad (6.24)$$

From these equations the invariant of the sum of the normal stresses must be verified:

$$\sigma_x + \sigma_y = \sigma_r + \sigma_\theta$$

The biharmonic equation is valid independently of the coordinates system; then the operator ∇^2 becomes from rectangular to polar coordinates:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = \frac{\partial^2}{\partial r^2} + \frac{\partial}{r \partial r} + \frac{\partial^2}{r^2 \partial \theta^2}$$

and Eq. 6.18:

$$\left(\frac{\partial^2}{\partial r^2} + \frac{\partial}{r \partial r} + \frac{\partial^2}{r^2 \partial \theta^2} \right) (\sigma_r + \sigma_\theta) = 0 \tag{6.25}$$

In order to satisfy equilibrium equations, the relationships between stresses and the new Airy function $\phi(r, \theta)$ must be the following (see details of the steps in [13]):

$$\begin{cases} \sigma_r = \frac{1}{r} \frac{\partial \phi}{\partial r} + \frac{\partial^2 \phi}{r^2 \partial \theta^2} \\ \sigma_\theta = \frac{\partial^2 \phi}{\partial r^2} \\ \tau_{r\theta} = \frac{1}{r^2} \frac{\partial \phi}{\partial \theta} - \frac{1}{r} \frac{\partial^2 \phi}{\partial r \partial \theta} = \frac{\partial}{\partial r} \left(\frac{1}{r} \frac{\partial \phi}{\partial \theta} \right) \end{cases} \tag{6.26}$$

Substituting Eq. 6.26 in Eq. 6.25 the biharmonic nature of ϕ is verified:

$$\left(\frac{\partial^2}{\partial r^2} + \frac{\partial}{r \partial r} + \frac{\partial^2}{r^2 \partial \theta^2} \right) \left(\frac{\partial^2 \phi}{\partial r^2} + \frac{\partial \phi}{r \partial r} + \frac{\partial^2 \phi}{r^2 \partial \theta^2} \right) = \nabla^4 \phi = 0 \tag{6.27}$$

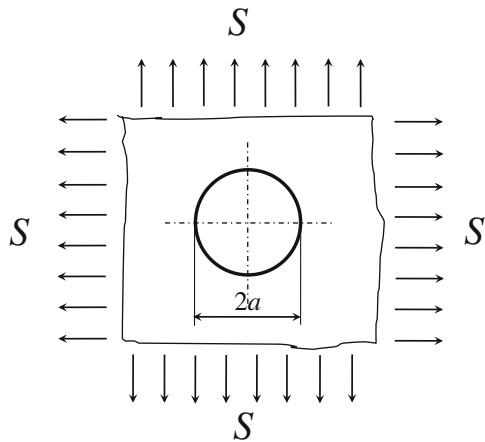
It is now possible to specialize the previous general expressions for some practical cases.

6.3.1.1 Uniform Load in Every Direction

For an axial-symmetric problem Fig. 6.13, the dependence is only on the variable r and the Airy function is function of r only:

$$\phi = \phi(r)$$

Fig. 6.13 Stress in a plate with a circular hole, uniformly loaded



Equation 6.27 reduces to:

$$\left(\frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr}\right) \left(\frac{d^2\phi}{dr^2} + \frac{1}{r} \frac{d\phi}{dr}\right) = 0$$

This is an ordinary differential equation that admits the following solution:

$$\phi(r) = A \ln r + Br^2 \ln r + Cr^2 + D \tag{6.28}$$

where A, B, C, D are constants to be determined by boundary conditions. Substituting Eq. 6.28 in Eq. 6.26 the following stress values are derived:

$$\begin{cases} \sigma_r = \frac{A}{r^2} + B(1 + 2 \ln r) + 2C \\ \sigma_\theta = -\frac{A}{r^2} + B(3 + 2 \ln r) + 2C \\ \tau_{r\theta} = 0 \end{cases} \tag{6.29}$$

For $r \rightarrow \infty$, being the stress S a finite value, B must be zero. The other boundary conditions are:

$$\begin{cases} \sigma_r = S & \text{for } r \rightarrow \infty \\ \sigma_r = 0 & \text{for } r = a \end{cases}$$

then:

$$\begin{cases} 2C = S \\ A = -S a^2 \end{cases}$$

and finally Eq. 6.29 give:

$$\begin{cases} \sigma_r = S \left(1 - \frac{a}{r^2}\right) \\ \sigma_\theta = S \left(1 + \frac{a}{r^2}\right) \end{cases} \tag{6.30}$$

Exercise 6.3 (*Stress concentration for a circular hole in a spherical thin vessel of large diameter, internally Pressurized*) (a) If the hole and the thickness are small respect to the sphere diameter in order that it can be considered in membrane regime, the previous theory can be extended to the case of a spherical vessel internally pressurized with a circular hole.

(b) The concentration factor due to a hole in a plane membrane in uniform state of stress is estimated through Eq. 6.30. For $r = a$ stress components are:

$$\begin{cases} \sigma_r = 0 \\ \sigma_\theta = 2S \end{cases} \tag{6.31}$$

Then the stress concentration factor, i.e. ratio between maximum circumferential stress and uniform stress, is $K_t = 2$.

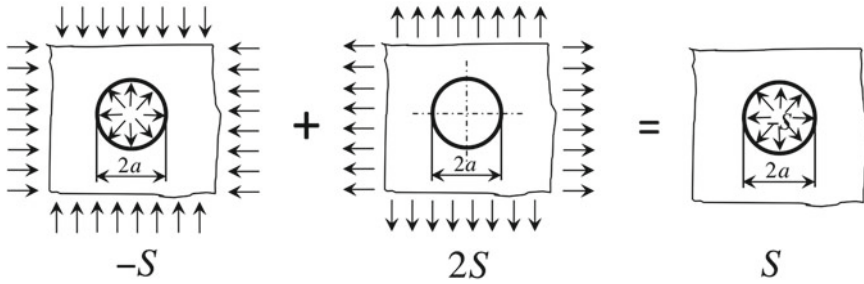


Fig. 6.14 Stress in a plate with a circular hole loaded on its boundary

Exercise 6.4 (*Stress concentration for a hole loaded on its boundary, in a large thin plate*) (a) This case is the over-position of the previous state plus a uniform state of stress, Fig. 6.14. The circumferential stress around the hole is given by:

$$\sigma_{\theta} = -S + 2S = S$$

(b) The circumferential stress due to a pressure S acting on the hole boundary is a tension stress with the same absolute value of the applied pressure.

6.3.1.2 Load in One Direction

The problem solving proceeds according to the following steps:

- It is assumed as first step, Fig. 6.15, that the disturbance of a small hole in a large plate uniformly loaded in one direction is localized in a limited area of the plate and practically absent at a large distance compared with the hole radius. For the Saint Venant principle [6], stress distribution on a radius b of a concentric circle,

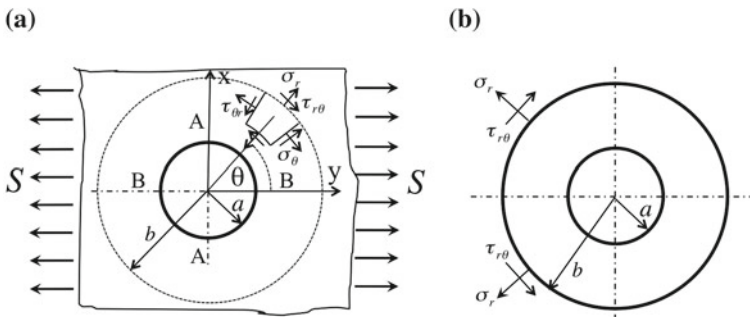


Fig. 6.15 a Stress in a plate with a circular hole loaded in one direction, b annular region of the plate

at a certain distance from the hole center, is essentially the same as in the plate without the hole.

- A tentative expression for the stress distribution on this circle of radius b is proposed for equilibrating the stress S applied in one direction at a large distance:

$$\begin{cases} \sigma_r = S \\ \sigma_\theta = 0 \\ \tau_{r\theta} = 0 \end{cases} \quad \text{for } \theta = 0 \text{ and } r \rightarrow \infty$$

The expression is:

$$\begin{cases} \sigma_r|_{r=b} = S \cos^2 \theta = \frac{S}{2}(1 + \cos 2\theta) \\ \tau_{r\theta}|_{r=b} = -\frac{S}{2} \sin 2\theta \end{cases} \quad (6.32)$$

- This stress state applied on the circle b is an over-position of the previous case of uniform stress in every direction plus a cosinusoidal distribution; then the stresses in the annular region of radii a and b can be regarded as composed of two parts: the first due to a uniform load $S/2$ and the remaining part due to a cosinusoidal stress distribution on the circle boundary:

$$\begin{cases} \frac{1}{2} S \cos 2\theta \\ -\frac{1}{2} S \sin 2\theta \end{cases}$$

- This second distribution admits a stress function of the form:

$$\phi = f(r) \cos 2\theta \quad (6.33)$$

Equation 6.33 must satisfy the compatibility condition $\nabla^4 = 0$ given by Eq. 6.27. Substituting the Eq. 6.33 in it, the condition generates the following ordinary differential equation, because r and θ are separate in two different functions:

$$\left(\frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} - \frac{4}{r^2} \right) \left(\frac{d^2 f}{dr^2} + \frac{1}{r} \frac{df}{dr} - \frac{4f}{r^2} \right) = 0 \quad (6.34)$$

with the general solution:

$$f(r) = Ar^2 + Br^4 + C \frac{1}{r^2} + D$$

and a stress function of this form:

$$\phi = \left(Ar^2 + Br^4 + C \frac{1}{r^2} + D \right) \cos 2\theta$$

From Eqs. 6.26 the following stress components derive:

$$\begin{cases} \sigma_r = - \left(2A + \frac{6C}{r^4} + \frac{4D}{r^2} \right) \cos 2\theta \\ \sigma_\theta = \left(2A + 12Br^2 + \frac{6C}{r^4} \right) \cos 2\theta \\ \tau_{r\theta} = \left(2A + 6Br^2 - \frac{6C}{r^4} - \frac{2D}{r^2} \right) \sin 2\theta \end{cases}$$

The integration constants are determined by the boundary conditions that are:

$$\begin{cases} \sigma_r|_{r=b} = - \left(2A + \frac{6C}{b^4} + \frac{4D}{b^2} \right) = S/2 \\ \sigma_r|_{r=a} = \left(2A + \frac{6C}{a^4} + \frac{4D}{a^2} \right) = 0 \\ \tau_{r\theta}|_{r=b} = \left(2A + 6Cb^2 - \frac{6C}{b^4} - \frac{2D}{b^2} \right) = -S/2 \\ \tau_{r\theta}|_{r=a} = 0 \end{cases}$$

- Assuming an infinitely large plate ($b \rightarrow \infty$ i.e. $a/b = 0$), the equations give:

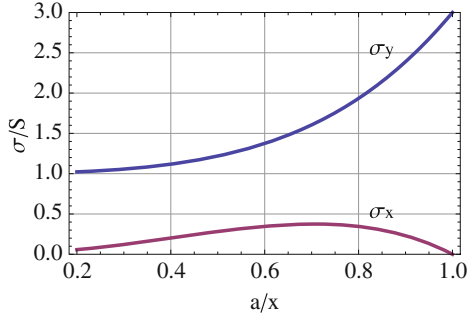
$$A = -\frac{S}{4} \quad B = 0 \quad C = -\frac{a^4}{4}S \quad D = \frac{a^2}{2}S \quad (6.35)$$

- The final expression of the stress components is, [6]:

$$\begin{cases} \sigma_r = \frac{S}{2} \left(1 - \frac{a^2}{r^2} \right) + \frac{S}{2} \left(1 - \frac{4a^2}{r^2} + \frac{3a^4}{r^4} \right) \cos 2\theta \\ \sigma_\theta = \frac{S}{2} \left(1 + \frac{a^2}{r^2} \right) - \frac{S}{2} \left(1 + \frac{3a^4}{r^4} \right) \cos 2\theta \\ \tau_{r\theta} = -\frac{S}{2} \left(1 - \frac{a^2}{r^2} \right) \left(1 + \frac{3a^2}{r^2} \right) \sin 2\theta \end{cases} \quad (6.36)$$

Along the x axis $A - A$ ($\theta = 0$) the principal stresses are, Fig. 6.16:

Fig. 6.16 Stress diagrams for σ_y and σ_x along the transversal symmetry axis x of the circular hole in an infinite medium, loaded in y direction



$$\begin{cases} \sigma_r = \sigma_x = \frac{S}{2} \left(1 - \frac{a^2}{x^2} \right) - \frac{S}{2} \left(1 - \frac{4a^2}{x^2} + \frac{3a^4}{x^4} \right) \\ \sigma_\theta = \sigma_y = \frac{S}{2} \left(1 + \frac{a^2}{x^2} \right) + \frac{S}{2} \left(1 + \frac{3a^4}{x^4} \right) \end{cases} \quad (6.37)$$

or:

$$\begin{cases} \frac{\sigma_x}{S} = \frac{1.5a^2}{x^2} - \frac{1.5a^4}{x^4} \\ \frac{\sigma_y}{S} = 1 + \frac{0.5a^2}{x^2} + \frac{1.5a^4}{x^4} \end{cases} \quad (6.38)$$

with:

- S = uniaxial stress at infinity in y direction
- σ_y = longitudinal stress
- σ_x = transversal stress
- x = distance from the hole centre
- a = hole radius

On the hole boundary σ_θ is obtained from Eq. 6.29 for $r = a$:

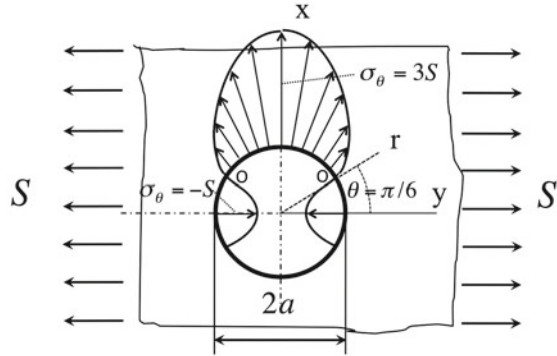
$$\begin{cases} \sigma_r = 0 \\ \sigma_\theta = S(1 - 2 \cos 2\theta) \\ \tau_{r\theta} = 0 \end{cases} \quad (6.39)$$

Zeros of the function σ_θ are $\theta = \pm \frac{\pi}{6}$. At these points O (Fig. 6.17) both principal stresses are null.³

Exercise 6.5 (Stress concentration due to a hole in a thin plate of a finite width, loaded in one direction) (a) The photoelastic model is a rectangular Araldite plate of width equal to 122 mm with a hole of 35.5 mm in diameter on the symmetry axis

³ In photoelastic theory points of this kind are called *singular isotropic points*, see Chap. 3 with isochromatics order equal zero.

Fig. 6.17 Polar stress diagram of σ_θ along the hole boundary



(Figs. 6.17 and 6.18a). The model was longitudinally loaded with 107.9 N (Newton) and stress state frozen. (b) The hypothesis of influence of external borders on the stress concentration in points A can be visualized by fitting the theoretical model for a hole in an infinite plate to the experimental data. The theoretical model for a single hole in a large plate (Fig. 6.18b) is given by Eq. 6.38:

$$\frac{(\sigma_y - \sigma_x)}{S} = 1 + \frac{0.5a^2}{x^2} + \frac{1.5a^4}{x^4} - \frac{1.5a^2}{x^2} + \frac{1.5a^4}{x^4} = 1 - \frac{a^2}{x^2} + \frac{3a^4}{x^4} \quad (6.40)$$

The derivative function of Eq. 6.33 is:

$$f'[x] = 1.25971 * 10^6/x^5 + 648/x^3 \quad (6.41)$$

Using this model is only an approximation, even if it fits very well to experimental data of this case.

(c) The approximation error can be estimated considering the real value of the maximum stress at hole edge, when the ratio $2a/H$ between the hole diameter and the width of the plate is not zero (that happens for H tending to infinity).

The equation for evaluating the maximum value of σ_y for finite values of $2a/H$ is the following:

$$\sigma_{max} = 0.284 + 2/(1 - 2a/H) - 0.600(1 - 2a/H) + 1.32(1 - 2a/H)^2 \quad (6.42)$$

and gives a maximum value on the hole boundary of 3.35 instead of 3 for infinite body.

(d) The stress can be expressed in MPa, remembering the relationship $N \cdot f_\sigma/d = 0.0871$ MPa, being the ratio $f_\sigma/d = 0.26/10$, for a photoelastic constant 0.26 N/ord.mm and thickness d equal to 10 mm. This value is valid for *Araldite B*, with the *frozen* method. Since the real value of the mean stress is 0.088 MPa (Load 107.91 N and cross-section of the bar equal to 122 cot 10 mm²), the approximation error with the use of the formula Eq. 6.40 valid for one hole in an infinite medium loaded in one direction is $\frac{0.088 - 0.0871}{0.088} 100 = 1.023 \%$.

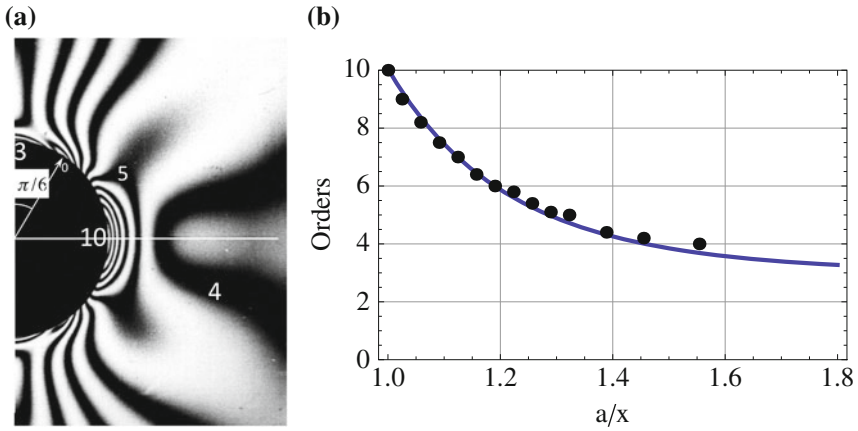


Fig. 6.18 a Fringes pattern with the diagram of $\sigma_1 - \sigma_2 \approx N_{order}$; b approximate fitting the theoretical model to data, Eq. 6.40

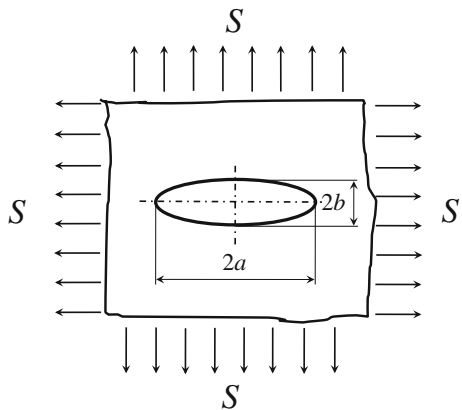
6.3.2 Elliptic Hole

Thin plate with elliptic hole is an appealing case in itself and also as introduction to the cracks study in Fracture Mechanics, Fig. 6.19.

For an elliptic hole with semi-diameters a and b , in a field of uniform tension, the problem is theoretically solved, introducing a system of elliptic coordinates [6]. The maximum stress is tangent to the hole perpendicularly to the major diameter: Eq. 6.43:

$$\sigma_{max} = 2\sigma_n \frac{a}{b} \tag{6.43}$$

Fig. 6.19 Elliptic hole in a thin plate loaded in two directions



In the case of uniform tension in a direction perpendicular to the main diameter the maximum stress at the border is:

$$\sigma_{max} = \sigma_n \left(1 + \frac{2a}{b} \right) \tag{6.44}$$

which, for $a = b$, gives a concentration factor of 3, as predicted by the previous case for a circular hole.

6.3.3 Concentration Factors for Slender Beams

For slender beams, concentration factors can be defined respectively for normal, bending and torsion stress, as [8]:

$$K_t = \frac{\sigma_{max}}{S} \quad \text{where : } S = \frac{F}{A} \quad \text{or : } S = \frac{M_f}{W} \tag{6.45}$$

$$K_t = \frac{\tau_{max}}{T} \quad \text{where : } T = \frac{M_t}{W_p}$$

with the following meaning for the symbols (referred to the notched cross-section):

- σ_{max} = maximum stress (by the theory of elasticity)
- S = nominal stress (by Saint Venant theory in the notched section)

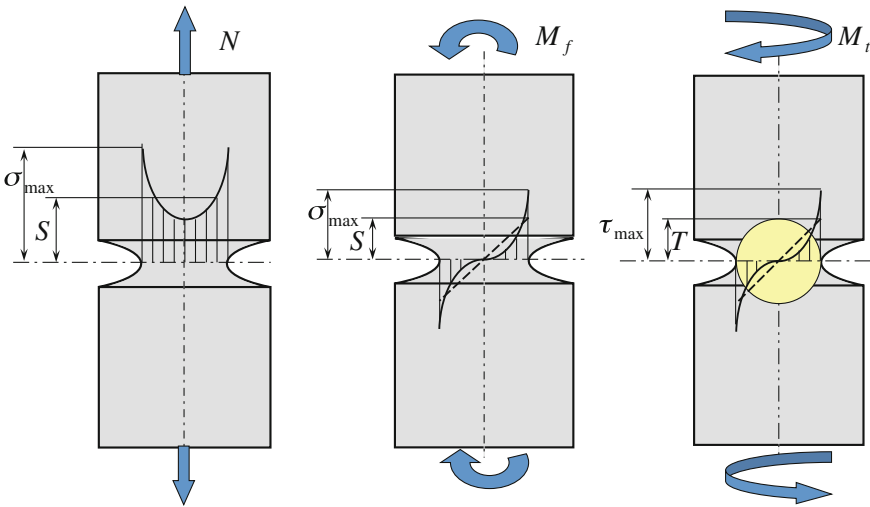


Fig. 6.20 Stress states at the root of the notch in slender beams in pure tension, bending and torsion

F = applied load
 M_f = applied bending moment
 W = bending section modulus of inertia
 M_t = applied torsion moment
 W_p = polar section modulus of inertia
 τ = maximum shear stress

Both stresses obtained by the theory of elasticity and by the elementary Saint Venant theory must fulfill the equilibrium relationships with external loads. In the first case of Fig. 6.20, the equilibrium condition is:

$$\int \sigma_t(r) dA = \sigma_{nom} \cdot A_k. \quad (6.46)$$

6.4 Stress Beyond the Elastic Limit

Beyond the elastic limit, proportionality between stress and strain no longer is verified. Strain state at the points of high concentration can be measured by optical or strain gage techniques as long as the analysis is limited to the surface. For high concentrations, strain gages must be sufficiently small, in order to avoid an underestimation of strains caused by the average measurement on the grid size,⁴ see Chap. 2.

The difficulty consists in determining the corresponding stress, because it needs the knowledge of the constitutive material law, i.e. of the relationship between strain and stress.

6.4.1 Imposed Displacements and Imposed Forces

Beyond the elastic limit it is necessary to distinguish two opposite situations that arise in the material when the structural element is loaded by external loads or subjected to external displacements.

When the element is a specimen of uniform section, stress is defined, in engineering hypothesis, as the ratio between uniaxial load and the cross-section area and strain as ratio between elongation and initial length of a reference specimen segment, and they are related one another by the so-called constitutive material law that, for metals as steel alloys or for other structural materials, has the characteristic look of Fig. 6.21a.

A change of this behavior must be expected not only when the specimen has other kind of loading, but also, for the same kind of load, when notches or grooves or geometric discontinuities are present. It is necessary to distinguish two limit cases that can be explained by the following opposite examples:

⁴ Producers offer very small gages, suitable for these applications, with longitudinal grid dimensions equal to 0.008 in. = 0.2 mm and resistance of $60 \div 120 \Omega$.

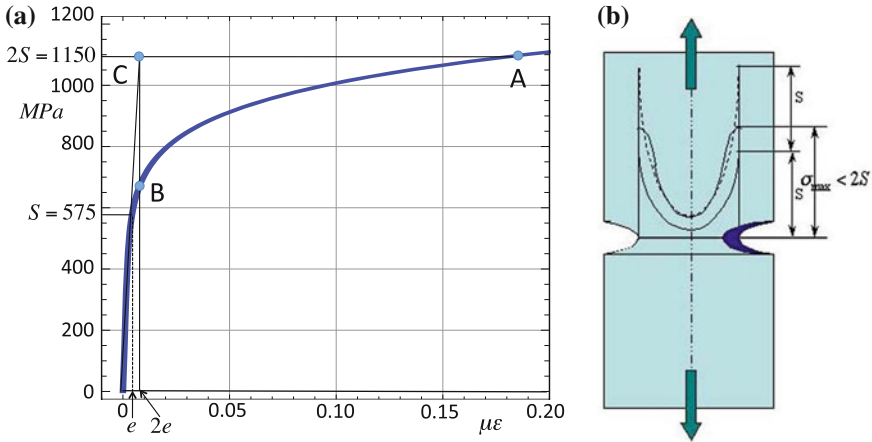


Fig. 6.21 a Diagram of constitutive law for a structural ductile material (NiCrMoV Alloy Steel), with the indication of the two extreme cases of an imposed load i.e. stress S (point A) and an imposed displacement, i.e. strain e (point B) and in b the intermediate most common case of a bar with a notch where stresses do not grow proportionally to load in all the points of the cross-section at the root of the notch

- For a thin pressurized tube the membrane stress is equal to $S = \frac{pr}{h}$, directly proportional to applied pressure, independent of the material behavior, because it derives from a pure equilibrium condition. If internal pressure grows, e.g., from p to $2p$, as long the material resists, the membrane stress S doubles ($2S$), as shown in the diagram of Fig. 6.21, and static state of the tube is not represented by the point C, but by the point A.
- If a thin tube, initially loaded at elastic limit with stress S and strain e is internally deformed by a very rigid punch, up to reach a final strain of e.g. $2e$, the representative point in the diagram, at the double value of initial strain e , is represented by the point B.
- In the most realistic case of a discontinuity between two different geometries as in the transition zone in pressure vessels between e.g. cylinder and sphere and generally in all the concentration points of slender beams, the state of material is intermediate respect to both cases seen before.
While for linear elastic behavior this distinction has no meaning, beyond the elastic limit three different models must be considered:
- In the case of imposed load (stress), the representative point in the stress-strain diagram is A, where the stresses are the same of an elastic problem ($2S$) but the strains are much greater.
- In the case of imposed deformation (strain), the representative point is B, where the deformations are the same of the elastic problem ($2e$) with much lower stresses.
- The most common case given by geometric discontinuities, as, e.g. in notched area of a shaft, Fig. 6.21b. According to the case shown in the figure, stress is

less than the first case and strain is greater than the second case. Only a small portion of material overcomes the elastic limit and yields, while the large volume of surrounding material remains in elastic field. The elastic volume creates thus a constraint on the smallest volume, imposing elastic displacements to its boundary.

As result, the stress state in the small volume is an intermediate situation, represented by a point that must be between *A* and *B* but, for the previous observation, closer to an imposed deformation, i.e. closer to point *B* than to point *A*, as the point *P* in Fig. 6.22b.

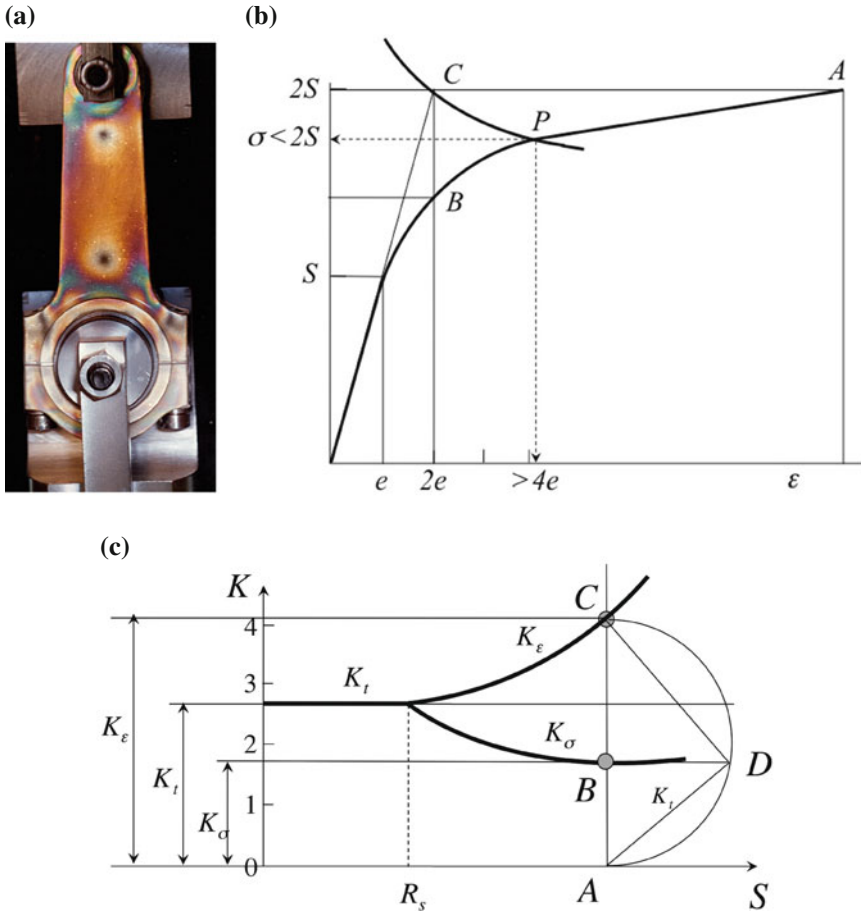


Fig. 6.22 Example of determination of the strain concentration at the *eye* of a connecting rod by reflection photoelasticity

(It is also obvious that, for equilibrium reasons, in the notch section a stress redistribution must take place, with a transfer of stresses from the most loaded fibers to adjacent less loaded fibers that give a *support effect*).

It can be concluded that locally, the stress state in geometric discontinuities is controlled more by deformation than by stress, even if the applied load keeps the main part of the body volume in elastic range.

6.4.2 Models for the Simulation of Constitutive Materials Laws

Starting from the case of cylindrical specimens with uniaxial loads, the search for a phenomenological model of material behavior is the first task for experimental stress analysis.

A model that can preserve its validity for the whole field of elastic and plastic behavior is preferable to two different models that describe separately elastic and plastic contributions.

The model that is commonly used is due to *Ramberg-Osgood*, which was originally developed for uniaxial state of stress but that maintains its validity for multi-axial states as equivalent stress in the most simple case of in-phase three-dimensional stress state and also for special biaxial loading condition as pure torsion, see a next Section:

$$\varepsilon_t = \varepsilon_e + \varepsilon_p = \left(\frac{\sigma}{E}\right) + \left(\frac{\sigma}{K}\right)^{1/n} \quad (6.47)$$

with this symbols meaning:

ε_t = total deformation

ε_e = elastic deformation

ε_p = plastic deformation

E = elastic modulus

K = coefficient of static plasticity

n = static plastic exponent

One of the advantage of this model is that fitting to experimental data can be done through a simple linear regression in the logarithmic scale. Today this point is no more so important for the capability of computer fitting procedures. Nevertheless, the classic way is the following:

- A load test on a uniaxial specimen of constant cross-section is performed at growing load, and stress is recorded, while strain is measured through electrical strain gages, photoelastic coating or other more recent and accurate optical techniques, in order to build a complete stress-strain relationship.
- From the measured (total) strain, the elastic deformation contribution, estimated with the relationship $\varepsilon_e = \frac{\sigma}{E}$ is subtracted (Table 6.2).

Table 6.2 Plastic term on the total strain for increasing values of applied stress in a NiCrMo steel alloy

σ (MPa)	ε_e	ε_p	ε_t
0	0	0	0
10	$4.8 \cdot 10^{-5}$	$4.7 \cdot 10^{-16}$	$= \varepsilon_e = 4.8 \cdot 10^{-5}$
100	$4.8 \cdot 10^{-4}$	$6.58 \cdot 10^{-9}$	$\sim \varepsilon_e = 4.8 \cdot 10^{-4}$
500	$2.3 \cdot 10^{-3}$	$6.4 \cdot 10^{-4}$	$2.9 \cdot 10^{-3}$
1000	$4.8 \cdot 10^{-3}$	$9.0 \cdot 10^{-2}$	$\sim \varepsilon_p = 9.5 \cdot 10^{-2}$

The residual term:

$$\varepsilon_p = \varepsilon_t - \frac{\sigma}{E}$$

represents the plastic deformation. This contribution is present only if the yield point is overcome.

The plastic deformation term is fitted by a model as the following one:

$$\sigma = K \varepsilon_p^n \tag{6.48}$$

- With this procedure it is possible to derive from the measured values of the plastic deformations at several load levels, the coefficient of static plasticity K and the plastic exponent n with a linear regression in logarithmic coordinates, Eq. 6.48. From the previous expression the inverse of plastic deformation is:

$$\varepsilon_p = \left(\frac{\sigma}{K}\right)^{1/n} \tag{6.49}$$

This term is combined with the elastic term in a single Equation that describes the whole field of elastic and plastic behavior of the material.

The binomial structure of the formula, only apparently seems unsuitable for describing the pure elastic behavior, for the supposed contamination of the plastic term, but actually it works very well also for the mathematical simulation of the pure elastic behavior, as it can be proved by this simple example.

Exercise 6.6 (*Plastic contribution on simulation of the elastic material behavior*)

(a) Given an uniaxial specimen of steel alloy with:

$$E = 210,000 \text{ MPa}$$

$$K = 1400 \text{ MPa}$$

$$n = 0.14$$

(b) Determine the elastic and plastic contributions on the total strain for increasing values of the applied stress.

(c) It is clear, from the Table 6.2, that the binomial expression is suitable to describe the material elastic-plastic behavior because the nonlinear term does not contaminate

the linearity of the first part of the curve and the same happens for the elastic term in the second part.

Exercise 6.7 (*Fitting to the experimental data a direct non linear equation, Eq. 6.47*)

(a) With the same data, determine the K and n values with a direct fitting procedure, Chap. 1.

(b) The following values are obtained by a direct fitting through an interpolation algorithm, to the same data of the Table in the same Exercise 6.6.

The results are: $K = 1381.6 \text{ MPa}$ $n = 0.134$. Such values slightly differ from the previous ones obtained by the canonical procedure ($K = 1400 \text{ MPa}$ $n = 0.14$).

6.4.3 Experimental Verification of the Neuber Model

The definitions of stress concentration factor and of strain concentration factor are respectively:

$$K_\sigma = \frac{\sigma_{max}}{S} \quad K_\varepsilon = \frac{\varepsilon_{max}}{e} \quad (6.50)$$

where σ and ε and are the local stress and strain in the discontinuity and S and e are the corresponding nominal values obtained by elementary beam theory. The link between e and S is given by the proportionality law $e = S/E$. In the elastic range so they have the same value.

Beyond the elastic limit, for the non-proportionality between local stress and local strain, concentration factors for stress and strain have not identical values but both differ from the stress concentration factor defined in elastic field for the same geometry.

The example refers to the uniaxial circumferential stress at the end of a connecting rod (statically loaded) in Fig. 6.22. Once it is measured the strain by strain gages located at the root of the notch or by photoelastic coating method (or by more accurate Digital Image Correlation method), Chap. 5, the strain concentration factor is given by $K_\varepsilon = \varepsilon/e$.

The corresponding stress concentration factor $K_\sigma = \sigma/S$ can be obtained if the constitutive law i.e. the stress-strain curve is experimentally determined and a corresponding model (as the previous Ramberg-Osgood model) is identified.

If the loading process is shown on a local stress-strain diagram, Fig. 6.22b, the elastic stress remains proportional until the point of coordinate e , S is reached. As it was observed, doubling, e.g., the applied load (the nominal stress), the corresponding point on the diagram is not $C(2e, 2S)$ because material state is not represented by the point A that is characteristic for an imposed local stress, nor by point B that would be the right one for imposed local deformation. The point representative of the real situation is an intermediate point P with a stress $\sigma < 2S$ and a strain $\varepsilon > 2e$. In other words:

$$K_\sigma < K_t < K_\varepsilon \quad (6.51)$$

A simple model due to *Neuber*, which is proven useful for simulating stress and strain behavior at discontinuities, is Eq. 6.51:

$$K_\sigma \cdot K_\epsilon = K_t^2 \tag{6.52}$$

This formula comes from studies on slender beams, Neuber [14–16] later extended to other cases of biaxial state of stress and to cyclic stress, i.e. to low cycle fatigue, where it is a basic simulation model of the nucleation process of small cracks, Chap. 8.

Figure 6.22c shows the trend of the three factors when the nominal stress S grows. Having considered an increment of this nominal stress, let B and C be the corresponding values of K_σ and K_ϵ at a certain level of the nominal stress.

For a theorem of elementary geometry it is possible to graphically verify the Neuber formula for those points. The circumference of diameter $A - C$ circumscribes the rectangular triangle whose side AD is the mean square of the two segments AB and AC i.e. it corresponds to the value of the stress/strain concentration factor in elastic field, K_t .

In this example of Fig. 6.22c K_ϵ is >4 (about 4.1) and $K_\sigma/\sigma/S$ is < 2 (about 1.7). The stress concentration factor in elastic field is 2.6 and the Neuber equation in this case is verified.

Substituting in Neuber equation the concentration factors formulas, we have:

$$\frac{\sigma_{max}}{S} \cdot \frac{\epsilon_{max}}{e} = K_t^2 \tag{6.53}$$

If the nominal stress remains in the elastic range, the following holds:

$$\sigma_{max} \cdot \epsilon_{max} = SK_t \cdot eK_t = \frac{(SK_t)^2}{E} \tag{6.54}$$

In Fig. 6.22b, Eq. 6.54 represents a hyperbola called *Neuber hyperbola* that passes through the point C of coordinates SK_t and eK_t . The point P is the common point of the stress—strain curve and the hyperbola and Eq. 6.52:

$$\begin{cases} \epsilon_t = \epsilon_e + \epsilon_p = \left(\frac{\sigma}{E}\right) + \left(\frac{\sigma}{K}\right)^{1/n} \\ \epsilon \cdot \sigma = \frac{K_t^2 S^2}{E} \end{cases} \tag{6.55}$$

make it possible to determine stress and strain at the root of the notch for a given external load.

Exercise 6.8 (*Model of stress and strain behavior at the root of a notch beyond the elastic limit*) (a) Determine stress and strain at the root of a notch in a structural element with a stress concentration factor, in elastic field, equal to $K_t = 2$ ($E = 210,000$ MPa, $K = 1400$ MPa and $n = 0.14$), loaded with a static nominal stress $S = 720$ MPa.

(b) Linear-elasticity dominates at all the points far from the notch. If linear elasticity is present also at the root of the notch, the following relationships are valid: $\sigma = K_t S = 3 \cdot 720 = 2160 \text{ MPa}$ and $\varepsilon = K_t S/E = 3 \cdot 720/210,000 = 1.03 \cdot 10^{-2} \text{ mm/mm} = 1.03 \% = 10,300 \mu\varepsilon$

In a diagram of the same kind of Fig. 6.22 σ and ε are the coordinates of the point C in the corresponding diagram. This is not the actual stress state: at the root of the notch, represented by the point P , is:

$$\varepsilon \cdot \sigma = 22.22$$

and a development of Eq. 6.55 gives:

$$\frac{22.22}{\sigma} - \frac{\sigma}{2.1 \cdot 10^5} - \left(\frac{\sigma}{1400} \right)^{1/n} = 0 \quad (6.56)$$

Zeros of the Eq. 6.56, are found by the use of iterative techniques (for example, the Newton's method@Newton).

In the present case a value $\sigma = 825 \text{ MPa}$ is found (instead of 2,160 MPa valid for the elastic solution). Vice versa, strain is much higher than in elastic field:

$$\varepsilon = \frac{22.22}{\sigma} = 2.69 \% = 26,900 \mu\varepsilon$$

instead 10,300 $\mu\varepsilon$.

The example shows the total inapplicability of the elastic solution when the yield point is locally overcome.

6.4.4 How Stress Modifies Material Behavior

If e.g., a slender cylindrical specimen loaded along its axis comes to failure by increasing monotonically the load, a ductile material response can change its behavior for various reasons that will be briefly examined, limiting the analysis to the effects induced by the stress state. Yield stress and ultimate stress for a specimens of uniform cross-section are assumed as characteristic data of a material, even if, as it is known, they vary in the case of different absolute dimensions as well as of different shapes (e.g. for unnotched and notched specimens), Fig. 6.23. The reason is due to three-axial stress states that arises at geometrical discontinuities, that modify the macro behavior of the specimen. At the yield stress and above, a notch tends to prevent the natural contraction of the root section: the yield and ultimate stresses thus increase with respect to the corresponding value of a standard specimen.

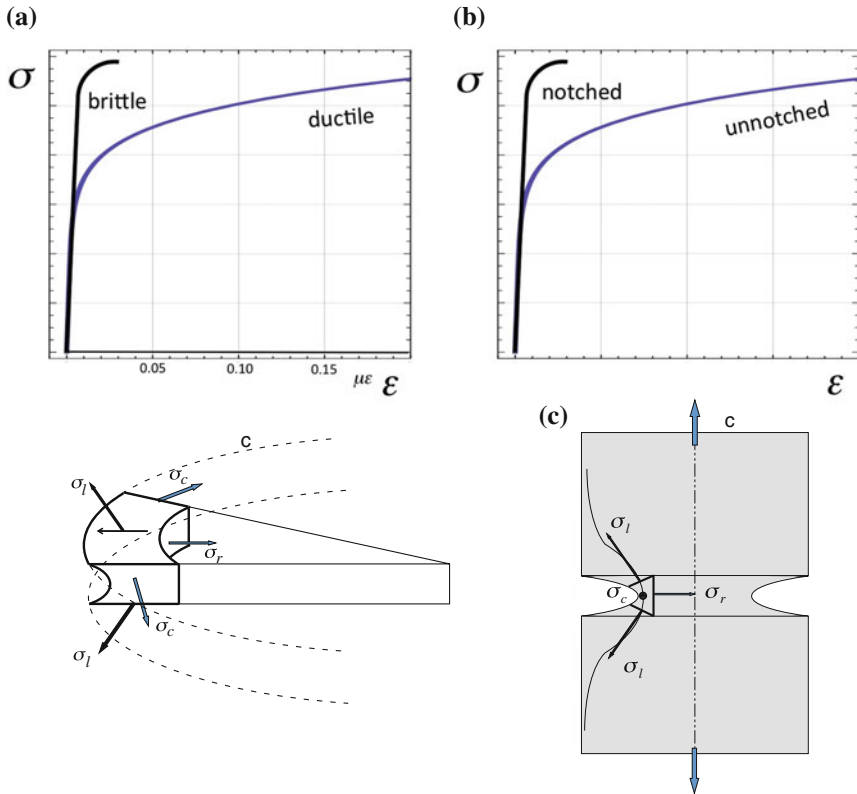


Fig. 6.23 a Stress strain curves for ductile and brittle materials and b influence of the specimen geometry (i.e. of stress distribution), on stress-strain curve, c local three-dimensional stress state

6.4.5 How Material Behavior Modifies Stress State

A cylindrical specimen of uniform cross section, when is close to failure, becomes locus of multi-axial stress state, with a variety of possibility of fracture surfaces.

In ductile materials, the stress state tends to become three-axial near the specimen axis, phenomenon that is caused by transverse contraction ($\nu = 0.5$) partially hindered by the most internal fibers. It generates brittle nucleation of a transversal crack close at the specimen axis, without appreciable plastic deformation, Fig. 6.24, after that the crack axi-symmetrically enlarges.

Moreover, the material layer on the symmetry cross-section of infinitesimal thickness is in plane strain state because it can't either shrink nor swell. As radial σ_r and circumferential σ_θ on the axis reach their maximum values, also the longitudinal stress along the axis should reach its maximum in order to preserve the condition of plane strain state. Both radial and circumferential stresses vanish on the external specimen surface.

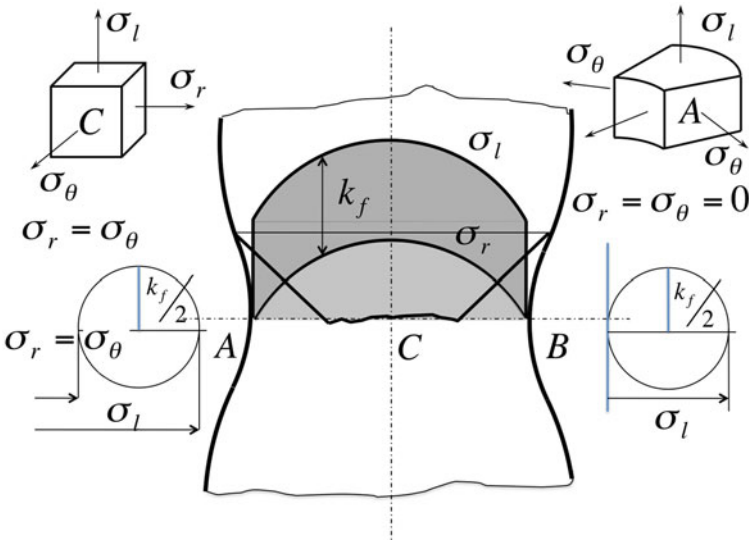


Fig. 6.24 Fracture surface for a cylindrical specimen of ductile material and explanation of the stress distribution on the mean plane

When the crack moves towards the external surface, the longitudinal stress σ_l dominates and necking of the cross section occurs, with a maximum shear stress at 45° that generates the typical conical-crater shape.

An elementary model is the following:

$$\varepsilon_l = \frac{1}{E} \cdot [\sigma_l - \nu(\sigma_r - \sigma_\theta)]$$

then:

$$\varepsilon_l = \frac{1}{E} \cdot [\sigma_l - 0,5 \cdot 2(\sigma_r)] = k_f/E$$

The Mohr circle which describes the stress state in points on the symmetry section, keeps constant diameter k_f , passing from the center to the outer surface.

6.4.6 Elementary Models of Physical Theory of Fracture

The origin of a micro-structural fracture can be identify in presence of defects at atomic lattice level which, under the stress action, evolve in macro-defects, which may eventually generate failure surfaces. A model that makes it possible the estimation of cohesion strength between two contiguous atomic planes and the calculation of the relative breaking stress, is developed, based on the knowledge of the experimental strain-stress law at atomic level, where stress is applied and strain follows as relative displacement of two adjacent lattice planes.

Starting from a regular lattice structure without any defect, it is possible to introduce two different de-cohesion models:

- Assuming the stress perpendicular to the lattice plane.
- Assuming the stress tangential to the lattice plane.

For both cases the models, on the base of experimental results, are similar. In both cases the relationship of the applied stress (normal in the first model and shear in the second model), can be partially represented by a sinusoidal curve.

6.4.6.1 Normal Stress Micro-structural Model

In the first case the model is shown in Fig. 6.25a. The stress required to push two atomic planes apart of the quantity x from their equilibrium position, with the assumption of a defects absence is, at a limited load level, reversible with a quasi-linear reaction due to the mutual atoms attraction. That is true until a maximum value is achieved. The straight line from the origin represents a initial behavior.

After this first portion, the diagram can be approximated by a sinusoidal function with a wave length λ :

$$\sigma = \sigma_{max} \sin \frac{2\pi x}{\lambda}$$

When the maximum is overcome, the attraction is reduced and finally vanishes. For small x values the curve is approximated by this law:

$$\sigma \approx \sigma_{max} \frac{2\pi x}{\lambda} \tag{6.57}$$

In the same interval the relationship between stress and strain can be considered linear and the concept of elastic modulus E can be introduced:

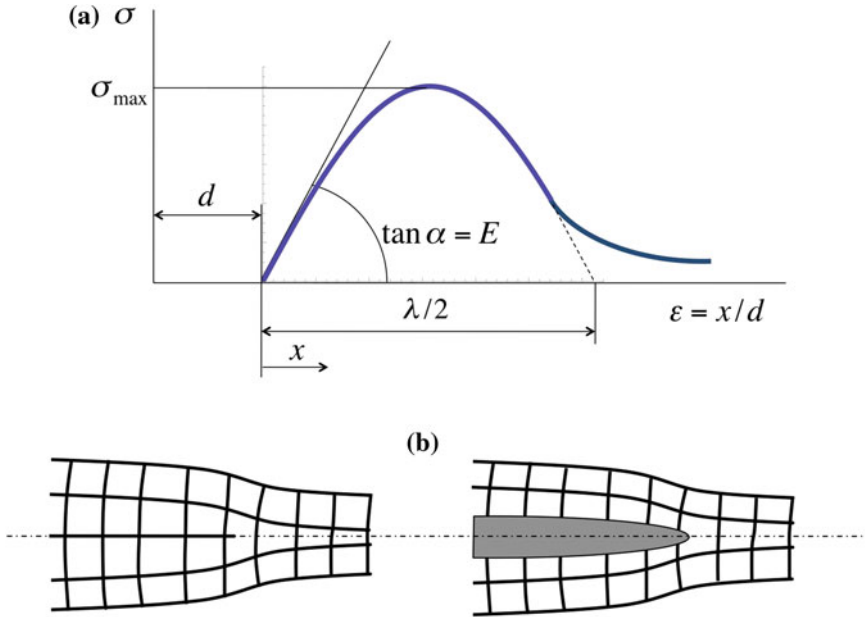


Fig. 6.25 a Stress-strain diagram between atomic planes for the first model and in b scheme of a dislocation and even a possible inclusion that concentrates the normal stress

$$\sigma = E \frac{x}{d}$$

where d is the distance between two lattice planes.

Since both expressions must be valid, an estimation of the maximum stress as function of λ , d and E is:

$$\sigma_{max} = \frac{E\lambda}{2\pi d} \tag{6.58}$$

This value can be interpreted as a theoretical estimation of the maximum tolerable stress value normal to atomic planes. Referring e.g. to steels, mean values for d and λ are:

$$d = 3 \cdot 10^{-7} \text{ mm} \quad \lambda = 2 \cdot 10^{-7} \text{ mm}$$

so that the maximum stress lies in the interval:

$$\frac{E}{15} \leq \sigma \leq \frac{E}{5} \tag{6.59}$$

much higher than all possible experimented value.

In fact, the crystals (or combinations of poly-crystals in grains), always, Fig. 6.25, contain defects (a) and inclusions (b). They lead to strong stress concentration values

between lattice bonds and drastically reduce the previously calculated stress of two orders of magnitude (e.g., for a microcrack of $3 \mu\text{m}$, equal to about $10^4 d$ stress on the tip of the crack can reach a value of: $\sigma_{tip} = 10^{-2} \sigma_{max}$).

6.4.6.2 Tangential Stress Micro-structural Model

Instead of normal stress perpendicular to atomic planes, a second mechanism of fracture due to shear stress along the atomic planes could be more realistic of the real material resistance.

The model is equivalent to the previous one but is developed for shear actions and x is in tangential direction:

$$\tau \approx \tau_{max} \frac{2\pi x}{\lambda} \quad (6.60)$$

In the first part of the curve the relationship between shear stress and shear strain can be considered linear, introducing the concept of shear modulus G :

$$\tau = G\gamma = G \frac{x}{d} \quad (6.61)$$

where d is the distance between two atomic planes and x is the displacement in τ direction.

Since both Eqs. 6.60 and 6.61 must be valid in the same interval, an estimation of the maximum shear stress as function of λ , d and G is approximately obtained:

$$\tau_{max} = \frac{G\lambda}{2\pi d} \simeq \frac{G}{2\pi} \quad (6.62)$$

In this case hypothesized defects are *dislocations* (plane, three dimensional or a combination of both), that have a direct role to favoring the generation of lattice slide planes, both in case of static as of cyclic loading.

This is due to a simplified mechanism of movement that needs much lower stress and energy level than the first model, as Fig. 6.26 tries to demonstrate: only one ligament at a time is necessary to solve for shifting the entire atoms block of one step. So dislocations movements and their stacking against the edges of the grains, with plastic deformation or eventually micro-cracks nucleation explain different behaviors of the material, plasticity and fracture. Only few attraction forces must be overcome in this case and this circumstance reduces drastically the value of the failure stress at macroscopic level.

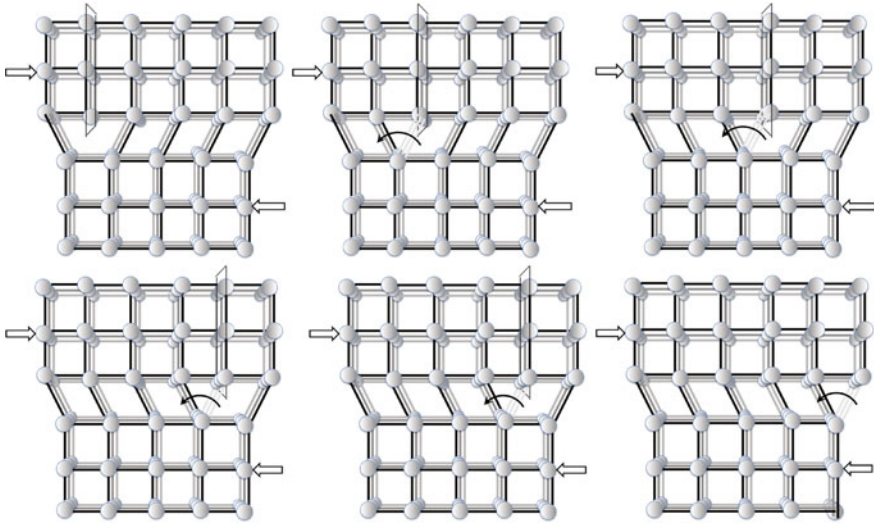


Fig. 6.26 Illustration of plane dislocation movement, until it reaches the grain boundary (sketch modified from technical literature). The movement of a plane dislocation needs a much lower effort than the one necessary to complete atomic planes to shear in the absence of dislocations

It can be concluded that the ductile and brittle behavior are not only result of chemical composition and treatments but are also a consequence of the stress state. It may happen that normally brittle materials can behave as ductile when they are loaded in a way to favor dislocation movements, e.g. with a high compressive component (i.e. with a triaxial compressive stress). In the same manner, normally ductile materials can increase their ductility if a compression is present and conversely reduce their ductility if a three-dimensional tensional stress (that inhibits the dislocation movements), is applied.

6.5 Special Equipment for Static Tests

Sometimes it is convenient to design and build special equipment for static tests, when e.g. the structural elements require very high loads that exceed the capacity of standard universal machines. That is the case e.g., of static tests on threaded connections for oil or gas pipes, Fig. 6.27.⁵ The equipment is designed for special

⁵ The equipment was developed for the Materials Development Center, Castel Romano (Rome) and there was implemented by a measurement system by a custom made strain gage load cell (Manufacturer: Giuliani-Forlì Italy).



Fig. 6.27 Pipes joint for $2\ 3/8'' \div 9\ 5/8''$ diameters

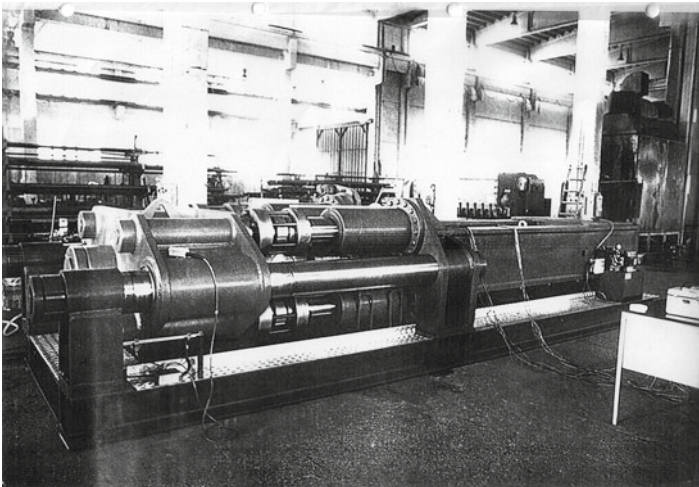


Fig. 6.28 Static test equipment for pipes joints for 13 MN in tension, 5 MN in compression [17]

loads combinations up to 13 MN in tension, 5 MN in compression and a bending moment up to $15^\circ/100'$ with tubes of $9\ 5/8''$ diameters, Fig. 6.28 [17].

An example of a larger apparatus, with an increased capacity of 25 MN in tension, is shown in Fig. 6.29. The design of these testing machines needs special care to avoid stress concentration points and secondary effects due to alignment tolerances that can induce high spurious bending. Moreover, when the specimen breaks, the surplus energy must be absorbed without damaging the machine. (Apparatus designed and built by “ITALSIGMA” in Forlì (Italy) under the authors supervision and calculation.)



Fig. 6.29 Static test equipment for pipes and joints of 25 MN capacity (manufacturer: Giuliani-Forlì, Italy)

References

1. Beer FP, Johnson ER Jr (1992) *Mechanics of materials*, vol 1, 2nd edn. McGraw-Hill, New York
2. Crandall SH, Dahl NC, Lardner TJ (1978) *An introduction to the mechanics of solids*, vol 1, 2nd edn. McGraw-Hill, New York
3. Lemaitre I, Chaboche JL (2000) *Mechanics of solid materials*, 4th edn. Cambridge University Press, Cambridge (original in French in 1985)
4. Timoshenko SP (1953) *History of strength of materials with a brief account of the theory of elasticity and theory of structures*, 1st edn. Dover Publications, New York
5. Rice JR (2010) *Solid mechanics*. http://esag.harvard.edu/rice/e0_Solid_Mechanics_94_10.pdf, School of Engineering and Applied Sciences, Harvard University Cambridge
6. Timoshenko SP, Goodier JN (1969) *Theory of elasticity*. McGraw-Hill, New York
7. Föppl L, Mönch E (1959) *Praktische Spannungsoptik*, vol 1, 2nd edn. Springer, Berlin
8. Pilkey WD, Pilkey DF (2008) *Peterson's stress concentration factors*, 3rd edn. Wiley, Hoboken
9. Freddi A (1972) Photoelastic stress analysis of a nuclear reactor pressure vessel. *Mecc J AIMETA* 7(3):183–204
10. Curioni S, Freddi A, Garro A (1971) Experimental and theoretical stress analysis in a nuclear steam generator head. In: Jaeger TA (ed) *Proceedings of the 1st international conference on structural mechanics in reactor technology*, vol 1. Berlin, Reactor Pressure Vessel
11. Freddi A, Dal Bianco A (1974) *Analisi fotoelastometrica delle tensioni in un contenitore di reattore nucleare soggetto a carichi esterni agenti sui bocchelli*. Quaderno CNR della Ricerca scientifica 85
12. Young W, Budynas R, Sadegh A (2011) *Roark's formulas for stress and strain*, vol 1, 8th edn. McGraw-Hill Professional. ISBN-10: 0071742476
13. Alfirevic I (2006) *Linear structural analysis. Thin-walled structures*. Pretei d.o.o. Zagreb
14. Neuber H (1946) *Theory of Notch Stress*, vol 1, 1st edn. J. W. Edwards, Ann Arbor
15. Neuber H (1958) *Kerbspannungslehre*, vol 1, 2nd edn. Springer, Berlin

16. Neuber H (1961) Theory of stress concentration for shear-strained prismatical bodies with arbitrary nonlinear stress-strain law. *Trans ASME J Appl Mech* 28:544–550
17. Curioni S, Freddi A, Podrini A (1992) A new 1000 tons machine for testing of tubular products and threaded connections in oil field. In: SEM (ed) *Proceedings of VII international congress on experimental mechanics*, society for experimental mechanics, SEM. <http://sem.org>

Chapter 7

Local Stress Models for Variable Loads

Abstract This chapter presents the conceptual basis of the phenomenological theory of fatigue at a high number of cycles; however, consistently with the approach of this book, this is only for establishing and validating local stress models, which are necessary steps to develop a theory as well as design criteria. The problem of local stress state in an elastic range is dealt with separately from the Low Cycle Fatigue controlled by local deformation. Although all these theories together with Fracture Mechanics have the task of explaining and simulating fatigue phenomena and are often presented together, they utilize different models for the description of the state of local stress, local strain and stress singularities. It therefore seems appropriate to consider them separately. After a reminder on the historical foundation of the discipline, the chapter first discusses the identification of the loads. Following a brief presentation of the classic models for handling the laboratory fatigue data, the relationship between stress concentration and fatigue parameters is examined. This is an inverse problem that is presented for slender and massive structures.

7.1 Stress Analysis for Developing a Theory on High Cycles Fatigue

Fatigue of materials is a term introduced in the first half of the nineteenth century to specify the phenomenon known as:

The process of progressive localized, permanent structural change, occurring in the material subjected to a condition that produces fluctuating stresses and strains at some point or points and that may culminate in cracks or complete fracture after a sufficient number of fluctuations [1, 2].

The empirical nature and the importance of the phenomenon, still today the main cause of structural failures, have led to the development of theories and models of phenomenological character.¹ The basic hypothesis for the fatigue at a high number of cycles is the elastic behavior of the material until failure.

¹ See extensive literature on Fatigue described and cited in classic books on Machine Design, Materials Science, Mechanics of Materials such as [2–5], and in reviews as e.g. [6] for specialized topics.

Historically, the real understanding of the fatigue phenomenon and the foundation of a theory that can overcome empirical rules required the development of two experimental facilities: special testing equipment in controlled environment that could reproduce in the laboratory the essential of the phenomenon as it happens in reality, and special devices and sensors for deformation measurement *in the field* and in the laboratory.

Practical phenomenological theories have been established supported by extensive research activities carried out over the years: these theories do not cover all applications, but they are useful for selecting a number of variables that are first responsible for a process essentially not predictable which is reflected in a large dispersion of data that can not be eliminated, even for tests in ideal conditions, as it happens in laboratories.

Today, the behavior of new materials and the influence of multi-axial types of load applied to bodies of complex shape are still intriguing problems on the research front.

The experimental approach on which every theory must be built primarily involves the following points:

- Availability of sensors, load cells, instruments and test equipments for laboratory and for in-field measurements.
- Measurement of the parameters for the characterization of material and geometry, which are intimately bound in constituting fatigue resistance.
- Statistical tools for dealing with random variables: data regularization from laboratory measurements (descriptive statistics) is necessary to transfer them to real situations (inferential statistics).
- Analysis of fracture surface: the appearance of fracture surfaces is in itself a carrier of information on the kind of loads and on the stress level.

7.2 Identification of Variable Loads

The measurement of *live* (or in-field) deformations and also in the laboratory, has allowed the identification of the variables responsible for fatigue. The germinal but complete work is due to Wöhler, [7], who developed a system for the measurement of flexural and torsional deformations on railway axles. Figure 7.1 shows the initial page of the first scientific document on fatigue. The method is very ingenuous and original and its principles are essentially unchanged today [8]:

- The device for measuring and recording the flexural and torsional deformations consists of leverages connected with the wheels that move an index with a sharp pin that, during rotation in real service conditions, leaves a trace of the wheel-plane deformation on a zinc plate. This deformation is proportional to the instantaneous axle bending that, in this way, is recorded, Fig. 7.2 on the left side.
- A similar device, based on a rigid tube coaxial to the axle, records the relative rotation between the two wheels mounted on the same axle, Fig. 7.2 on the right.

Ausbau des Kirchenschiffs die Summe von ... 1318612 Thlr. 21 Sgr. 1 Pf. Ueber die zur Baukasse eingezahlten Beträge Seitens des Central-Dombau-Vereins enthält sein Rechenschaftsbericht das

Nähere. Die Summe derselben beläuft sich auf 478200 Thlr., ohne die Beiträge der anderen Vereine von 118731 Thlr. Köln, den 18. Mai 1858.

Zwirner, Königl. Geheimer Regierungs- und Bau-Rath.

Bericht über die Versuche, welche auf der Königl. Niederschlesisch-Märkischen Eisenbahn mit Apparaten zum Messen der Biegung und Verdrehung von Eisenbahnwagen-Achsen während der Fahrt, angestellt wurden.

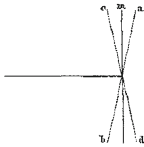
(Mit Zeichnungen auf Blatt D' und E' im Text.)

Beschreibung der Apparate.

Die Apparate sind auf beiliegenden Zeichnungen (Blatt D' und E' im Text) dargestellt. Es waren je ein Paar für Stahl-Achsen von 3 1/2 Zoll und für eiserne Achsen von 5 Zoll Durchmesser in der Nabe angefertigt worden.

Die Apparate zum Messen der Biegung der Achsen bestehen aus einem in der Mitte der Achse auf dieser befestigten Stücke A, das an seinem einen Ende einen Zapfen a trägt, welcher als Drabungs-Axe für zwei Zeiger dient, von denen der eine mittelst des Zwischenstücks bb, mit dem einen, der andere mittelst des Zwischenstücks cc, mit dem anderen Rade der Art verbunden ist, daß, wenn die Achse sich biegt, die Räder also ihre normale Lage verlassen, die zweiten Enden c der beiden Zeiger sich auf dem bis dahin verlängerten Stücke A verschieben. An den Enden e der Zeiger sind Reißerspitzen angebracht, welche auf Zinkplatten, die am Stücke A befestigt sind, die Größe der Verschiebung für jedes Rad besonders aufreißten.

Um die, einem bestimmten Zeiger-Ausschlag entsprechende Kraft festzustellen, wurden die Achsen mittelst Dynamometer gebogen, welche am Umfange der Räder angebracht waren und diese gegen einander zogen.



Da während der Fahrt die Punkte, an denen die Räder mit den Zeigern in Verbindung gesetzt sind, abwechselnd über und unter der Achse sich befinden, wird der Ausschlag des Reiffers nach zwei Seiten hin erfolgen, so, als wenn bei der ruhenden Achse das Rad sowohl die nebenstehend punktirte Lage cd als die ab angenommen hätte. Bei gleicher, auf Biegung wirkenden Kraft ist mithin während der Fahrt der Zeiger-Ausschlag doppelt so groß, als bei der ruhenden, mittelst Dynamometer gebogenen Achse.

Beide Apparate zum Messen der Biegung sind so construirt, daß 1 Zoll Zeiger-Ausschlag während der Fahrt einer Bewegung ac am Umfange des Rades von 1/4 Zoll oder einer Abweichung am von der normalen Lage von 1/8 Zoll entspricht.

Die Seitenkraft, welche am Umfange des Rades angebracht werden muß, um eine gleiche Biegung der Achse, also einen einseitigen Zeiger-Ausschlag von 1/4 Zoll hervorzubringen, ist für die Achsen von 3 1/2 Zoll Durchmesser in der Nabe, mit Rädern von 36 1/2 Zoll Durchmesser, = 23 1/2 Ctr., und für die Achsen von 5 Zoll Durchmesser in der Nabe, mit Rädern von 36 1/2 Zoll Durchmesser, = 70 1/2 Ctr.

Die Apparate zum Messen der Verdrehung der Achsen bestehen aus einem Blechcylinder mn (auf den Kupfertafeln), welcher sich leicht um die Achse drehen läßt; derselbe ist mit dem Arm f an das eine Rad O so befestigt, daß der ganze Cylinder, also auch der am andern Ende angebrachte Arm g, sich mit diesem Rade bewegt. Wenn also eine Verdrehung der Achse stattfindet, wird dieser Arm g sich gegen das andere Rad O, entsprechend verschieben. Derselbe trägt einen Zapfen h, um den sich ein Zeiger dreht, dessen eines Ende c mittelst eines Zwischenstücks mit dem Rade O, verbunden ist, so daß, wenn der Arm g sich verschiebt, das zweite, mit einem Reißer versehenen Zeiger-Ende k auf einer an den Arm befestigten Zinkplatte einen Riß macht, dessen Länge der Verschiebung und auch der Größe der Verdrehung der Achse proportional ist.

Die einem bestimmten Zeiger-Ausschlag entsprechende Torsionskraft wurde durch angebrachte Hebel und Gewichte gemessen; da aber während der Fahrt die Verdrehung der Achse so stattfindet, daß sowohl das eine als das andere Rad vorrückt, erfolgt der Zeiger-Ausschlag nach beiden Richtungen und ist mithin bei gleicher Kraftwirkung doppelt so groß, als bei der mittelst Hebel und Gewicht nur in einem Sinne hervorgerufenen Verdrehung.

Der Apparat an 3 1/2 Zoll Durchmesser ist so construirt, daß 1 Zoll Zeiger-Ausschlag einer Bewegung von 0,321 Zoll am Umfange des Rades von 36 1/2 Zoll Durchmesser entspricht; gegen die normale Lage des Rades beträgt also die Größe der Bogen-Abweichung 0,100 Zoll, oder der Torsionswinkel 30 Minuten.

Zu einer solchen Verdrehung ist eine am Umfange des Rades wirkende Kraft von 18 1/2 Ctr. erforderlich. Bei dem Apparat der Achsen von 5 Zoll Durchmesser in der Nabe, deren Räder 36 1/2 Zoll Durchmesser haben, ist auf 1 Zoll Zeiger-Ausschlag die Bewegung am Umfange des Rades = 0,228 Zoll, die Abweichung gegen die normale Lage also 0,114 Zoll, und der Torsionswinkel = 21 Minuten. Um eine solche Verdrehung hervorzubringen, ist eine am Umfange des Rades wirkende Kraft von 44 Ctr. erforderlich.

Ausführung der Versuche.

Die Achsen von gleichem Durchmesser mit den Vorrichtungen zum Messen der Biegung und zum Messen der Torsion wurden unter je einen Wagen gebracht, und diese dann dem Betriebe übergeben. Die Wagen wurden nach zurückgelegter Fahrstrecke nicht gedreht, so daß also abwechselnd jede der beiden Versuchs-Achsen Vorder- und Hinter-Ende war.

Die Versuche sind sowohl mit sechsrädrigen als mit vier-rädrigen Wagen ausgeführt; von letzteren konnten sowohl bedeckte als offene genommen werden. Die Wagen gingen, zwei Fälle ausgenommen, nur in Güterzügen.

Für jede Reise sind neue Zinkplatten am Reißer-Apparat 42*

Fig. 7.1 First page of the famous Wöhler paper of 1858 on strain measurements on railway axles

- These measurements, reproduced in the laboratory with static application of unknown forces and torsion couples applied to the wheels, allow the identification of the values of the real forces produced on the axles during service (This is a rare example of an experimental solution of an inverse problem, as described in Chap. 1).

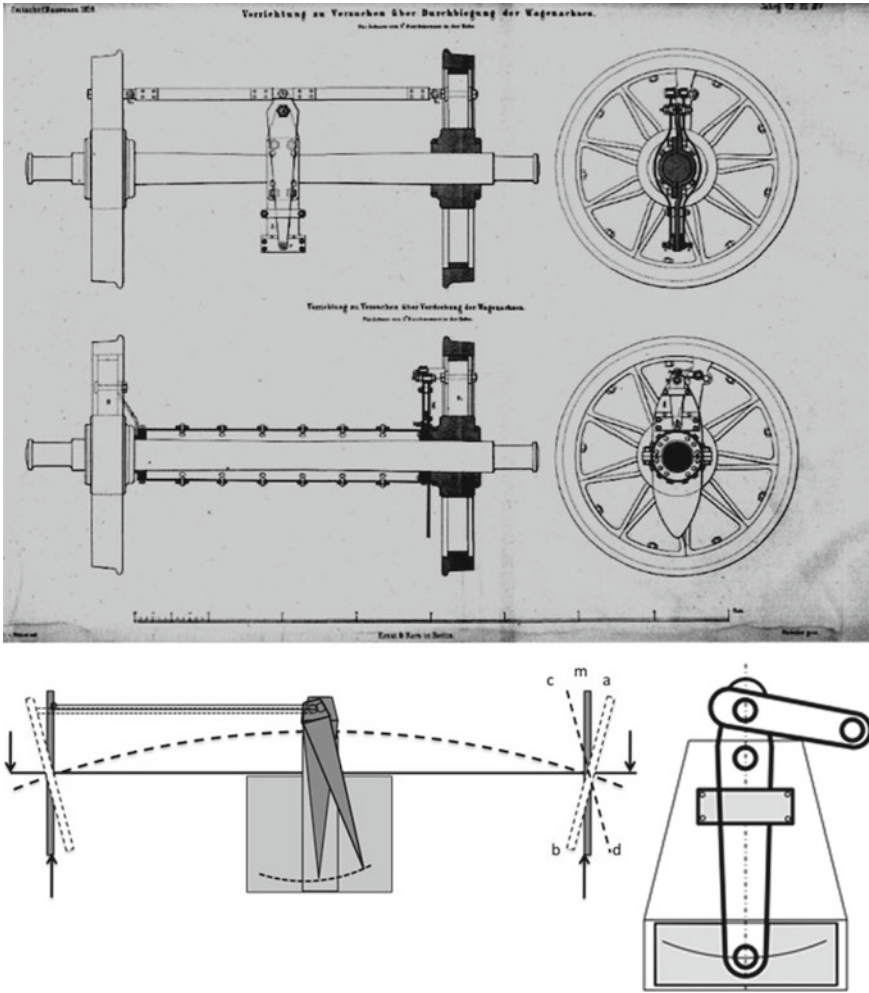


Fig. 7.2 Original Wöhler's drawing with the devices for measuring and recording flexural and torsional deformations of a railway axle. A sharp pin generates a scratch on a zinc plate, proportional to the instantaneous bending or to the torsional deformations

- In a second step, these identified forces are cyclically applied to specimens with the same size and shape as the axles and, from the results of the tests, a model is built that represents the life dependence of the axles on the amplitude of the loading cycles.



Fig. 7.3 Device to measure the six forces and moments transmitted by a ski boot to a ski

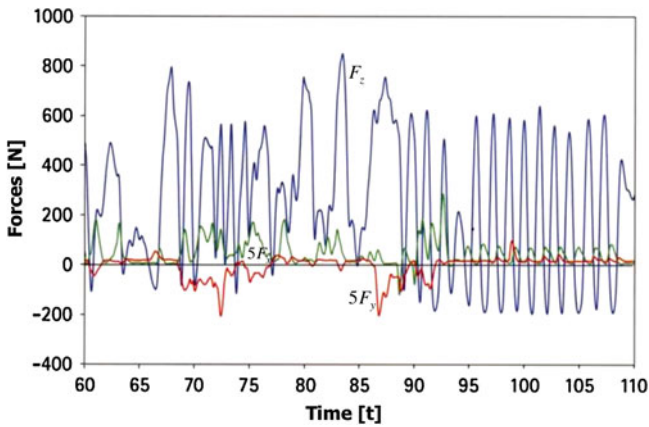


Fig. 7.4 Example of a recording of in-field loads: three forces F_x , F_y , F_z transmitted by a ski boot to a ski

7.2.1 Case Study of Fatigue Loads on a Sport Equipment

The previous method is applied to a case study with the same task of identifying *in-field* forces like the Wöhler problem. This is an example of an identification problem of three Forces and three Moments applied to a ski boot, Fig. 7.3a. The loading-cell has seven micro-cells, each of which is calibrated for a force measurement [9–11]. The problem is even-determined and well-posed. Figure 7.4 shows an example of recording of the three loads.

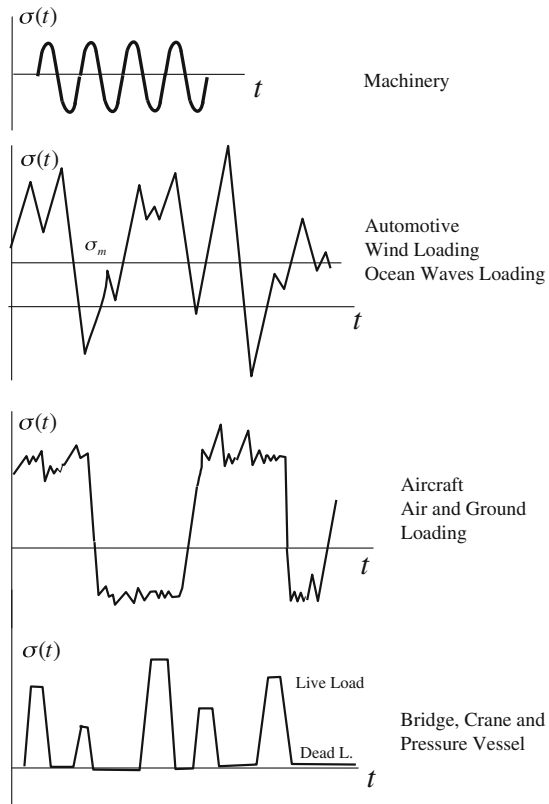
The solution of this design problem is discussed in Chap. 2, Sects. 2.10.2.1 and 2.10.2.2 and related Exercises, Fig. 2.48. Other examples of load identifications in sport or industrial devices by *in-field* tests in operation conditions are reported in [13–18].

7.2.2 Types of Loads

Structural components are generally exposed to a complex, sometimes random, sequence of loads, of different amplitudes. In machinery, the history is in general of constant amplitude load. The automotive components are dominated by a significant tensile or compressive mean stress and are often random distributed, like wind loading and ocean waves. Suspension helical springs have variable shear stress plus a mean torsion shear stress. Vice versa, the vehicle bracket load history is essentially fully reversed loading with zero mean stress, like a cantilever beam deflected at the free edge and released to a free vibration. Aircraft have two different loading levels in the air and on the ground. A thin or thick-walled pressure vessel subjected to cyclic internal pressure represents a component subjected to mean tensile stresses, while bridges and cranes have compressive mean stress [19, 20].

Concluding, in fatigue phenomena fully reversed loads together with tensile and/or compressive static mean loads are very common in all fields of the engineering [21]. Figure 7.5 shows an elementary classification of loading types.

Fig. 7.5 Examples of different types of loading, modified from [12]



For High Cycle Fatigue, stresses are elastically determined. Spectra of local stress diagrams relating the stress levels to the number of cycles can be considered irrespectively of the order of application. In order to specify the load history acting on a component, it is thus enough to collect the parameters of the service stress spectra, for the amplitudes and the mean stress values.

This operation must distinguish the proportional or at least synchronous stress (with superposition of the same type of stress as normal stress and shear stress), from non-proportional stress case. The spectra parameters determination for the last case can become very complex and proper guidelines must be followed, such as [22].

Under variable amplitude loading, the definition of stress *cycles* can become ambiguous and it is preferable to define stress *reversals*. At constant amplitude, one cycle is made up of two reversals. This aspect will be discussed in Chap. 8, where the counting methods will also be presented.

From the point of view of stress analysis simulation models, it is only appropriate to remember that:

- If the sequence of load is random, a statistic distribution function must be found, hypothesizing it on the basis of experience.
- If, vice-versa, the sequence is deterministically known, a variable load history is reduced to a series of *equivalent* single-cycle loadings through a counting method such as the *rainflow* method (presented in Chap. 8). It is now possible to reduce the real sequence to a histogram of stress cycles, thus creating a fatigue spectrum.
- For each stress level, the degree of cumulative damage is determined from the corresponding S-N curves, defined together with their values of the mean stress.
- A combination of the single-cycle damage is done using the algorithm of linear damage cumulation, such as *Palmgren-Miner's* algorithm, see [22].
- Since structural components are often loaded in several directions while the fatigue testing is obtained by uniaxial loading conditions for the elementary state of internal forces such as tension/compression, bending and torsional moments, the equivalence rule is necessary. If proportional or at least synchronous stress states are present at the reference point, they can be superposed or, in the case of different directions, reduced to equivalent stresses.

7.3 Laboratory Testing

The Chief of the Harz mines, *Wilhelm August Julius Albert*, is considered as being the first to set up an experiment to test fatigue in chains. He developed a testing apparatus with a waterwheel Fig. 7.6a. In 1837, Albert reported the results of these experiments [23], establishing a correlation between loads and number of chain loading cycles.

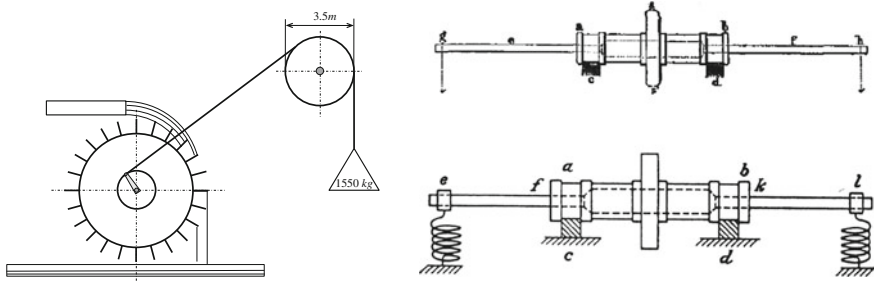


Fig. 7.6 Early testing equipments built to ascertain the phenomenon of fatigue by Albert (on the left) and by Wöhler (on the right) [24]

Nevertheless, the engineer *August Wöhler*, was the first to carry out a comprehensive research campaign in the fatigue area, from in-field measurements to methodical laboratory tests, showing the way to the development of all the subsequent design rules and of material qualification to prevent fatigue. Through a series of experiments conducted in the second half of the nineteenth century:

- He recognized the criticality of stress concentration points and the dependence of the effect of weakening of the axles on the geometry and material.
- He experimented with special self-made equipments, powered by a steam engine, Fig. 7.6b, and found the existence of a limit amplitude of the applied stress.
- He determined the detrimental dependence of the amplitude of fatigue limit as the mean positive (tension) stress increases.
- He developed a special testing machine to apply fatigue tests in torsion.

7.3.1 Rotating Bending Test

The specimen, generally with an hourglass shape to facilitate rupture in the central cross-section, is mounted on two oscillating supports with the constraint of two hinges, Fig. 7.7.

Two equal forces applied to the collars, generate a constant bending moment along the specimen. Taking movement from an electric motor, the system generates a rotating bending, since each fiber is cyclically subjected in variable time to a sinusoidal stress variation, as shown in the same figure. From the point of view of the stress analysis, the only necessary control concerns the measurement of the applied force and the counting of loading cycles until specimen failure.

The symbols have the following meaning, Fig. 7.7:

- σ_o upper stress
- σ_u lower stress
- $R = \sigma_u / \sigma_o$ stress ratio

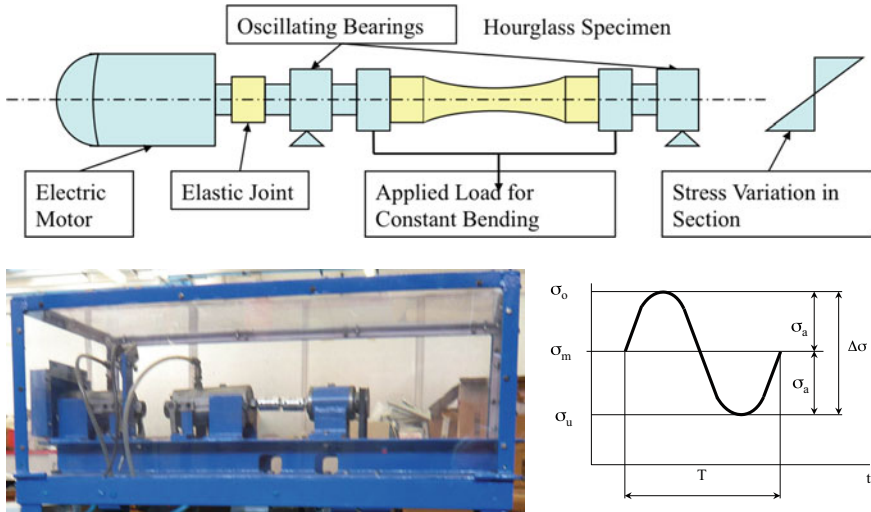


Fig. 7.7 Scheme of a rotating bending machine and stress variation in the central cross-section of the specimen

- $\sigma_m = \frac{\sigma_o + \sigma_u}{2}$ mean stress
- $\sigma_a = \frac{\sigma_o - \sigma_u}{2}$ amplitude
- $\Delta\sigma = \sigma_o - \sigma_u$ range
- $f =$ frequency

$R = -1$ called *fully reversed* condition, is the first reference test, while $R = 0$ is another common reference test, called *pulsating tension*.

7.3.2 Tension/Compression Test and Bending Test in One Plane

There are other classic modes of fatigue testing with the specimen loaded axially with uniform stress distribution or in bending in one plane only. They can be carried out with a resonance machine, Fig. 7.8 (a) [25] or with servo-hydraulic or electro-mechanical machines. With an added device, (b) a bending moment can be applied alternately on a plane containing the axis of the machine.

The load cell that employs electrical strain gages must have high stiffness, Chap. 2, Sect. 2.8.

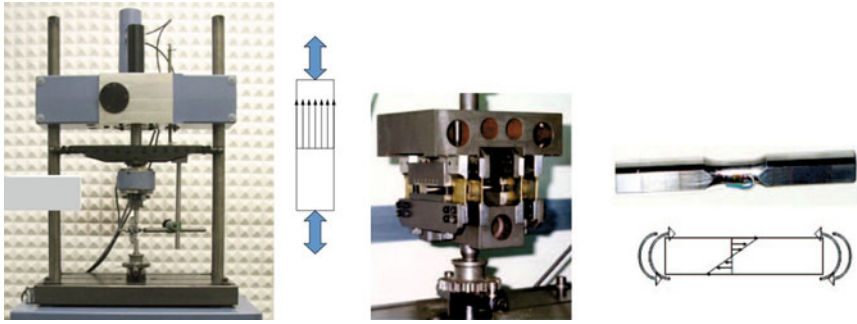


Fig. 7.8 Push/pull test or bending test in a resonance machine, with the schemes of the specimens and a special device for generating bending in one plane

7.3.3 Torsion Test

This test too can be performed with a resonance machine with a special device that transforms push/pull into pure torsion [26] or with a servo-hydraulic universal machine, both equipped with a mechanism that converts axial load into torsion moment, Fig. 8.14 in Chap. 8, that shows the functional principle of the method.

In Fig. 7.9 the specimen is instrumented with two strain gages at 45° from the axis, in order to control the deformations transmitted by the system at the resonance condition. This is a suggested precaution to make sure of the torsional load actually transmitted.



Fig. 7.9 Instrumented test specimen for alternate torsion and related equipment to be mounted on a vibrofore [26]

7.4 Fatigue Data Processing

To understand how a state of cyclic stress influences material behavior, the phenomenon must be isolated by applying a stress state associated with one type of internal force (e.g. tension/compression, or rotating bending or alternate torsion), in smooth specimens, with a very high degree of surface finishing in order to prevent unwanted variability on the results. Every test is repeated with different values of the amplitude, recording the cycle numbers corresponding to the complete failure of the specimen.

7.4.1 Fatigue Data Collection

The pairs of values (amplitude of the stress cycle versus number of cycles to failure) are reported in a cartesian diagram:

- In the ordinate, the stress amplitudes (in the middle cross-section of the specimen) σ_a in the case of tension/compression or completely reversed bending moment (or similarly τ_a in the case of completely reversed torsion moment). Both are calculated with the Saint Venant theory from the knowledge of the applied loads.
- In the abscissa, the corresponding number of cycles to failure.

The simulation models are obtained in the form of interpolating curves, plotted in logarithmic coordinate on one or on both axes. In the first case, data compression for the logarithmic scale is achieved on the horizontal axis only, and the curve shows a tendency to an asymptote. In the second case, the experimental points tend to order themselves in straight lines, with considerable simplification and help to the simulation.

Figure 7.10 shows a comparison between the curves for steels in semi-logarithmic coordinates Fig. 7.10a and in double logarithmic coordinates Fig. 7.10b. The first segment is an uncertain zone, despite some models developed for dealing with this region. The reason can be found in the different behavior of a failure in this zone that is dominated by a high stress level, which corresponds to a local plastic deformation.

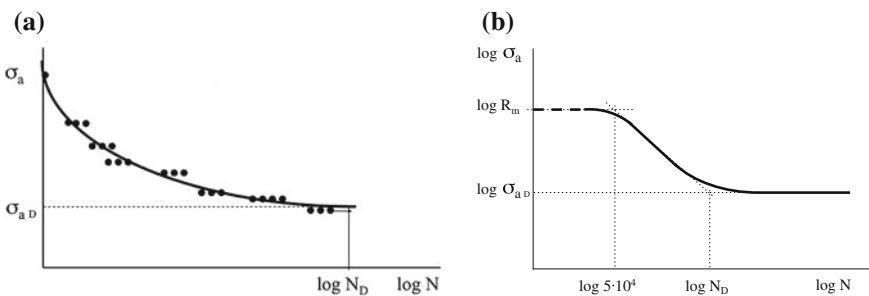


Fig. 7.10 Representations of the fatigue curve in semi-logarithmic and double logarithmic coordinates with their asymptotes

Nevertheless, at least for some kind of materials, an elastic behavior in this area too cannot be excluded.

As reference, material curve tests must be performed with a stress ratio $R = \sigma_{min}/\sigma_{max}$ ratio (between the minimum and maximum cycle stress) equal to -1 , for bending or for tension/compression loadings. The hourglass shape of the specimen with an increment of stress of about 5 % in the central cross-section, with respect to a cylindrical specimen with constant cross-section, slightly alters the results, but it is sometimes preferable to ensure that the failure is generated in the central area, far from the grip zones, whose fracture would invalidate the test.

- Fatigue damage is a phenomenon of stochastic nature which must be interpreted statistically. The distribution of the values of the cycles to failure, for the same stress amplitude applied, is supposed with a normal distribution for the logarithms of the coordinates. The variance is greater than that typical of static behavior and the coefficient of variation (ratio between the standard deviation and the mean) varies for the sloped segment, from 0.25 to 1.50, with mean values of about 0.4–0.6.
- The number of cycles to failure increases as the amplitude of the applied stress decreases.
- Below a certain amplitude value, the failure likelihood decreases strongly, to justify, for practical engineering use, the definition of a threshold, called *endurance fatigue limit* (for infinite lifetime).
- Fatigue failure occurs macroscopically like a *brittle fracture*, not accompanied by plastic deformation and, vice-versa, like a phenomenon of progressive modification of material structure, with damage accumulation for each loading cycle.
- The structural modification is generated where the stresses are highest, generally in *the surface layers*, then proceeding inward.
- A fatigue test different from the rotating bending test (e.g. in alternating bending or in tension/compression) has qualitatively a similar trend, but the characteristic values are different. In other words, the *Stress distribution* influences the fatigue behavior and the duration of the test.
- Frequency of stress cycle has little influence if overheating of the specimen and the effect of the environment, (e.g. corrosion) are avoided. Vice-versa, a synergic action stress/environment takes place and the frequency of load assumes a significant role.
- The mean stress has a strong influence on fatigue life and must be simulated in fatigue modeling. Sometimes it is the result of a residual stress state that can be originally present or sometimes introduced in a controlled way, to increment fatigue life.
- The enucleation of fatigue fracture is a highly *localized* phenomenon that is manifested only in a *limited volume* of stressed material with macroscopic and/or microscopic geometrical discontinuities.

- The surface state has an important role in fatigue resistance. A great deal of research has been developed on this topic. The difficulty to achieve complete simulation confines this point to the use of tables and diagrams that give coefficients to modify the resistance parameters. For this reason the material fatigue tests must be performed in strictly controlled surface conditions and the results can, with care, be transferred to real cases.

7.4.2 Model for Fatigue Behavior

The presentation is limited to the fundamental parameters of fatigue at a High Number of Cycles.

With reference to Fig. 7.10 the fatigue curve is interpreted in semi- and in double logarithmic coordinates. In the last case, three fields are discriminated:

- fatigue at a very low number of cycles, conventionally in first approximation, represented by a horizontal segment, $N < 5 \cdot 10^4$
- limited life-time at a finite number of cycles, represented by an oblique segment

$$5 \cdot 10^4 < N < 2 \cdot 10^6$$

- σ_{aD} resistance to infinite life (endurance fatigue limit), represented by a horizontal asymptote $N > 2 \cdot 10^6$

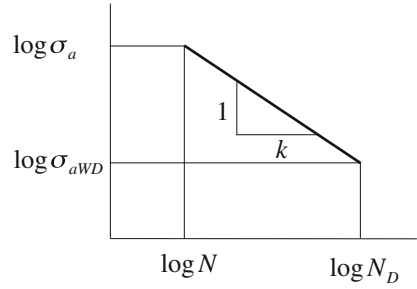
In the curve (b) of Fig. 7.10 the symbols have the following meanings:

- R_m static resistance
- k slope of the limited resistance
- N_D number of cycles at the knee.

Parameters k , σ_{aD} , N_D describe the whole fatigue behavior in addition to R_m , the value to which the curve tends for a very small number of load cycles.

The first horizontal segment can enter the elastic analysis only if stress remains under the elastic limit, that is a rare situation for ductile materials at a low number of cycles (and high stress level). This explains why a second theory of fatigue is developed, Chap. 8, based on the analysis of local deformations, particularly suitable for stress states beyond the elastic limit. Some steels have a quite pronounced endurance fatigue limit while others, such as e.g. aluminum alloys, do not show an asymptote parallel to the x axis. Recent experiences show that even steels can fail for a number of cycles much higher than those generally regarded as limits, (for $N \geq 10^9 \div 10^{10}$). The endurance fatigue limit, a conventional parameter, requires a statistical method to be determined, Sect. 7.13. Given a point with coordinates σ_{aD} and N_D , corresponding to the beginning of the horizontal segment, the slope of the

Fig. 7.11 Slope of the fatigue curve in log-log scale



inclined portion can be expressed with a constant k positive definite, as the ratio (Fig. 7.11):

$$k = (\log N_D - \log N) / (\log \sigma_a - \log \sigma_{aD}) \tag{7.1}$$

or:

$$\log \frac{N_D}{N} = k \log \left(\frac{\sigma_a}{\sigma_{aD}} \right)$$

i.e.:

$$\frac{N_D}{N} = \left(\frac{\sigma_a}{\sigma_{aD}} \right)^k$$

or:

$$N \cdot \sigma_a^k = N_D \cdot \sigma_{aD}^k = A_0 \tag{7.2}$$

In logarithmic coordinates (Basquin model) we have:

$$\log N + k \log \sigma_a = \log (N_D \cdot \sigma_{aD}^k)$$

i.e.:

$$x + k \cdot y = a \tag{7.3}$$

with: $x = \log N$ $y = \log \sigma_a$ $a = \log (N_{aD} \cdot \sigma_{aD}^k)$

The constant k has typical values in the interval from about 3 for specimens with geometrical discontinuities, to 5 for beams in bending or to 12 for beams in torsion, [27, 28].²

² The smaller values of k correspond to a more pronounced slope, while for higher values of k the oblique portion tends to become horizontal.

7.5 Models for Notch Effect

For slender structures with geometrical discontinuities (e.g. a cylindrical body with circumferential groove) subject to cyclic loading with ($R = -1$), the Notch Fatigue Factor K_f (in German literature β_k) is defined for each basic stress state (push/pull, bending, torsion) as:

$$K_f = \beta_k = \frac{\sigma_{aD}}{\sigma_{aDk}} \tag{7.4}$$

where σ_{aD} is the *nominal* endurance fatigue limit of a *smooth specimen* with the diameter of the notch root, and: σ_{aDk} is the *nominal* endurance fatigue limit of the *notched* reference specimen. The semi-logarithmic diagram in Fig. 7.12 is an example [29], for push/pull tested specimens. It might be argued that notched specimens have fatigue limits of smooth specimens reduced by the factor K_t .

The dotted curve at the bottom of the diagram (a) shows that this estimation does not agree with the experimental results (triangular marks) of the actual notched specimens.

The most realistic estimation (especially beyond $5 \cdot 10^4$ cycles, i.e. in actual high cycle fatigue condition) of the endurance fatigue limit for notched specimens,

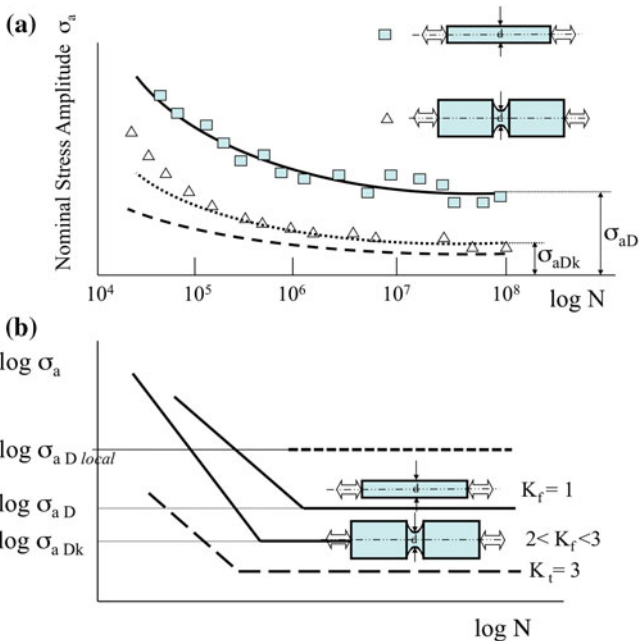


Fig. 7.12 a Semi-logarithmic and b double-logarithmic fatigue curves for notched specimens, obtained from the fatigue curve for smooth specimens reduced by the stress concentration factor K_t and by the fatigue factor K_f

is obtained from the fatigue curve of the corresponding smooth specimens, with a reduction K_f , being $2 < K_f < K_t$. In this example $K_t = 3$. The experimental measurements for fatigue at a low number of cycles do not agree with this prevision, as it is possible to observe in Fig. 7.12a.

The fatigue factor (defined for the loading condition $R = -1$) interprets the reduction of the fatigue life in a notched specimen compared to a smooth specimen. This conclusion depends on material properties, as well as on the concentration factor K_t but also on the kind of load (tension/compression, bending, torsion).

7.5.1 Models for the Micro-plasticity Zone

In order to clarify the difference between K_f and K_t , it is necessary to look at the physics of the material at the root of the notch.

The crack nucleation can be explained by the dislocations theory, Chap.6, Sect. 6.4.6.2, Fig. 6.26. This explains the displacements of the atomic planes placed in the most favorable directions of shear stresses at 45° with respect to the maximum normal stress (that is tangent to the surface). This slip movement generates not totally reversible extrusions and intrusions from which small plastic valleys and peaks and, progressively, micro-cracks arise after a sufficient number of cycles. The formation of the plastic slip is confined to a small but finite volume of material, Fig. 7.13.

This volume is circumscribed by a curve at $\tau_{max} = const.$ i.e. the threshold value for a critical shear stress at the root of the notch, Fig. 7.14 and can thus be visualized by the isochromatic curves, see e.g. Fig. 7.18. The observations of the micro-cracks clarify the reason for a lower value of K_f in comparison with K_t and justify a hypothesis on the fatigue notch factor: fatigue damage is due not only to a single stress value but to the whole state of stress in a small volume, close to the maximum value. It is thus reasonable to consider thus the effective critical stress not as the maximum elastic value but as the mean value on a finite small volume: $\sigma_{eff} = K_f \cdot \sigma_{nom}$. A second definition of K_f , can thus be obtained as, Fig. 7.14:

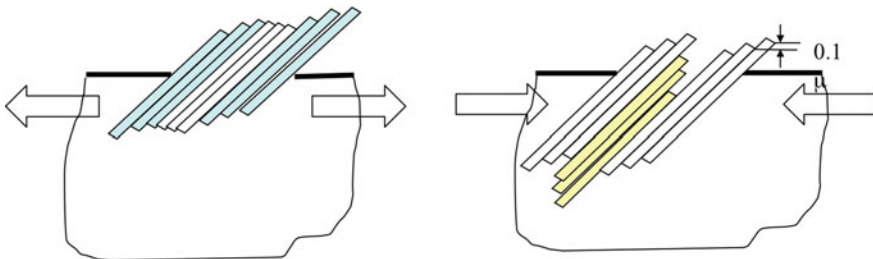
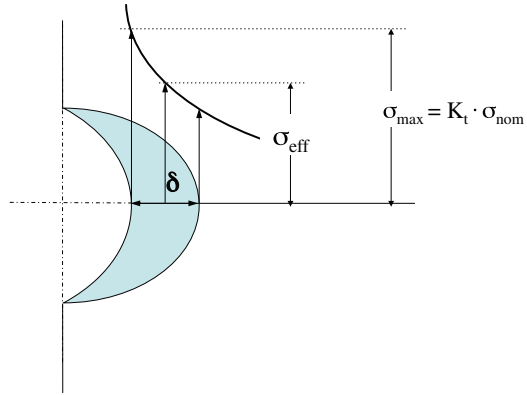


Fig. 7.13 Generation of extrusions and intrusions for a cyclic load at the surface of a body

Fig. 7.14 Shape of the maximum shear stress at the root of a notch



$$K_f = \frac{\sigma_{eff}}{\sigma_{nom}} < K_t = \frac{\sigma_{max}}{\sigma_{nom}} \tag{7.5}$$

The reduction of K_f is not to be interpreted as due to an extensive plasticity, that is not compatible with the fatigue phenomenon at a high number of cycles. The phenomenon remains essentially elastic, hypothesizing only a local effect of micro-plasticity.

7.6 Support Factor

In order to overcome the limits of the Saint Venant theory based on slender (1D rod-shaped) or thin (2D shell-shaped) components, a theory was developed for extending the integrity assessment to 3D block-shaped bodies. It is not unusual for some parts to be massive in mechanical systems or, more often, for some details in slender constructions to have three-dimensional states of stress. In these cases neither the elementary theory hypothesis is respected, nor a reference cross-section can be defined.

The theory introduces the concept of *Support Factor* at the most stressed points, sometimes called critical points or *reference points* [22] where a potential fatigue crack can nucleate. This hypothesis assumes that the fatigue strength is not a material property but a component property (at its reference points), so that the assessment of fatigue strength for a component is done comparing the state of stress with the fatigue limits at those points. The theory defines the strength, through the concept of the stress gradient, a local value that takes into account the support effect of the points close to the reference ones, i.e. of the effect of the confining material volume.

This is not the only theory that tries to evaluate the support of internal fibers. A proposal [30] based on an idea of Lazzarin et al. [31], argues that the local fatigue strength is sensitive to the integral of the stress contiguous to the most stressed points. This integral can be extended on a critical volume, on a critical area or on a critical length, depending on the geometry of the discontinuity.

The difficulty of this theory is the practical determination of these volumes, areas or length sizes. These parameters are determined by the value of ΔK_{th} , a Fracture Mechanics parameter. In this way a unified theory between fatigue and Fracture Mechanics is achieved [32]. In spite of this attractive perspective, the Fracture Mechanics parameter is not easy to determine, especially in the case of surface treatments and possible states of residual stresses in thin surface layers.

7.6.1 Support Factor in Slender Bodies

Starting from slender bodies for which the stress concentration factor as well as the fatigue factor can be defined, it is possible to write in a different way the previous Eq. 7.4, here repeated³:

$$\sigma_{aDk} = \frac{\sigma_{aD}}{K_f}$$

Multiplying both members for K_t , having called n the ratio:

$$n = \frac{K_t}{K_f} \quad (7.6)$$

the following equation is obtained:

$$\sigma_{aDk} \cdot K_t = \sigma_{aD} \cdot n \quad (7.7)$$

or:

$$n = \frac{\sigma_{aDk} \cdot K_t}{\sigma_{aD}} \quad (7.8)$$

In spite of the relationship between n and η , the n factor, called *Support Factor*, expresses a more general concept than η , because it keeps its validity also for not slender bodies.⁴ Before demonstrating this point, it is appropriate to remember two points:

³ The authors decided to adopt, for mathematical symbols, indexes and formulas, the notation used in the German Fatigue Handbooks (as [22]). It offers unambiguity in distinguishing among symbols, in spite of a high number of suffixes. The apparent difficulty of the choice avoids a number of doubts that arise in interpreting the meaning of a great number of variables.

⁴ This factor is related to the more common *Notch Sensitivity Factor* so defined:

$$\eta = \frac{K_f - 1}{K_t - 1} \quad (7.9)$$

by the equation:

$$n = \frac{K_t}{\eta(K_t - 1) + 1}. \quad (7.10)$$

- The stress concentration for each kind of internal stress state (push/pull, bending, torsion) is defined referring it to the nominal area of specimen cross-section, *cleansed* by the discontinuity.
- the Notch Fatigue Factor K_f , for each kind of internal stress state (push/pull, bending, torsion) is defined as:

$$K_f = \beta_k = \frac{\sigma_{aD}}{\sigma_{aDk}} \quad (7.11)$$

where σ_{aD} is the endurance fatigue limit (max. value for $R = -1$) of a smooth specimen with the diameter of the notch root and: σ_{aDk} is the endurance fatigue limit (max. value for $R = -1$) of the notched specimen. It is important to remark that both factors are conventionally defined, all the values of σ_a being *nominal values* and, conventionally, the endurance limit of a smooth specimen could be assumed, for each type of internal stress, for a tension/compression fatigue case.

7.7 From Slender to 3D-Bodies

Equation 7.8 can be read as follows (see Fig. 7.12):

$$n = \frac{\sigma_{aD \text{ local}}}{\sigma_{aD \text{ smooth}}} \quad (7.12)$$

where:

$$\sigma_{aD \text{ local}} = \sigma_{aDk} \cdot K_t \quad (7.13)$$

i.e. the n factor can be defined as:

The ratio between the (endurance) local limit at a reference point of the body (peak, not nominal, strength value) and the (endurance) limit of a smooth specimen of the same material having uniform stress distribution, assumed as reference.

The term:

$$\sigma_{aDk} \cdot K_t \quad (7.14)$$

has the meaning of a *Local Endurance Fatigue Limit* because it represents the material fatigue resistance at a given point. The maximum stress at that point must be compared to this value.

With this extension, the theory of the support factor is able to model three-dimensional bodies, remaining compatible with the simplified theory for slender bodies. Factors K_f , K_t and consequently n have an alternative meaning if they refer to material or to applied stresses:

- Referring to material: the n factor represents the **support** given to the fatigue resistance by the material close to it, due to the beneficial effect of local limited

plasticity that redistributes the strain on the nearest points.⁵ This is the reason for the name *Support Factor*.

- Referring to the applied stress: n represents the reduction of the actual stress with respect to the static one. In this meaning, the factor assumes, in some books, the alternative name of *Dynamic Stress Reduction Factor*.
- No support is offered to local strength in two cases:
 1. For absence of material plasticity.
 2. For uniform stress distribution on the cross-section.

7.7.1 Modeling Non-uniform Stress States by Means of the Relative Stress Gradient

The parameter that distinguishes a non-uniform stress distribution in a geometrical discontinuity from the uniform stress distribution in a reference specimen is the stress gradient, normalized with respect to the maximum stress at the reference point. It is thus the *Relative Stress Gradient* RSG, defined as:

$$\chi' = \frac{1}{\sigma} \frac{d\sigma}{dx} \quad (7.15)$$

Figure 7.15 explains the meaning of this parameter, keeping the maximum stress constant. When the stress gradient is high, Fig. 7.15a, the χ' is high and the volume affected by local shear deformation is small: the support effect is high, because the nearest fibers collaborate and give support to the most loaded ones. Figure 7.15b shows the other extreme case of low stress gradient, i.e. the low value for χ' . The volume affected by maximum stress is larger and support effect is minimum with a reasonable probability of micro-crack nucleation.

Moving from the beams theory to three dimensional bodies, instead of:

$$K_f = f(K_t, mat.)$$

the following relation has to be considered:

$$n = g(\chi', mat.) = \frac{\sigma_{aDk} \chi' \neq 0}{\sigma_{aD} \chi' = 0} \quad (7.16)$$

where χ' takes the place of K_t .

⁵ Satisfying the equilibrium of stresses over the section leads to a partial transfer of the load from the most loaded to the less loaded fibers, that is not possible when stresses are uniformly distributed in the cross-section.

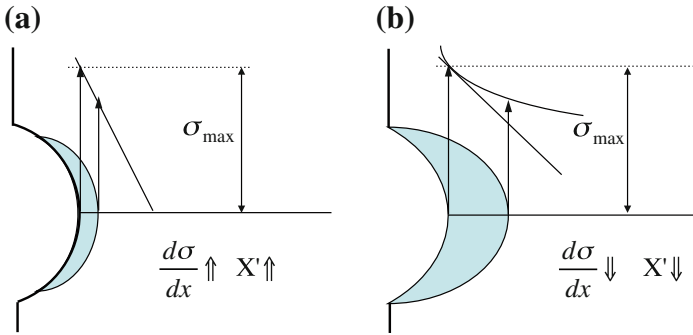


Fig. 7.15 The relative stress gradient χ' determines the dimensions of the most stressed material zone for a fixed value of σ_{max}

7.7.1.1 Relative Stress Gradient in Slender Beams

For simple shapes the χ' can be calculated, at least approximately, by the theory of elasticity. Equivalence theories like the Von Mises theory can handle the simplest cases of multi-axial states of stress (Fig. 7.16).

According to the definition of Eq. 7.15, the following expressions should be applied in the cases of push/pull and bending stresses:

	Assiale	A flessione	A torsione
	$\chi' = \frac{2}{\rho}$	-	-
	$\chi' = \frac{2}{\rho}$	$\chi' = \frac{2}{\rho} + \frac{2}{d}$	-
	$\chi' = \frac{2}{\rho}$	$\chi' = \frac{2}{\rho} + \frac{2}{d}$	$\chi' = \frac{2}{d} + \frac{1}{\rho}$
	$\chi' = \frac{2}{\rho}$	$\chi' = \frac{2}{D+d} + \frac{2}{\rho}$	$\chi' = \frac{2}{D+d} + \frac{1}{\rho}$
	-	-	$\chi' = \frac{2}{\rho}$

Fig. 7.16 Approximate theoretical expressions of RSG for common discontinuities in structural elements

$$X' = \frac{1}{\sigma_{max}} \cdot \frac{d\sigma}{dx} = 0 \qquad X' = \frac{1}{\sigma_{max}} \cdot \frac{d\sigma}{dx} = \frac{2}{d}$$

7.7.1.2 Example of Experimental Determination of the Relative Stress Gradient

An experimental method is based on the photoelastic analysis [33].

The Relative Stress Gradient can be written as:

$$\chi' = \frac{1}{\sigma_1} \frac{d\sigma_1}{ds_2} \tag{7.17}$$

where σ_1 is the principal stress along the isostatic s_1 , assuming a positive stress value, Fig. 7.17. The stress σ_1 is tangent to the boundary in a vertical direction and s_1 is the isostatic of the first family. Since $\sigma_2 = 0$, Eq. 7.17 gives:

$$\chi' = \frac{1}{\sigma_1 - \sigma_2} \frac{d(\sigma_1 - \sigma_2)}{ds_2} + \frac{1}{\sigma_1 - \sigma_2} \frac{d\sigma_2}{ds_2}$$

Remembering the equilibrium equations according to *Maxwell*:

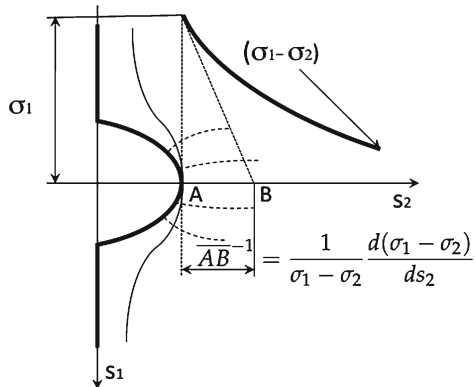
$$\frac{d\sigma_2}{ds_2} + \frac{\sigma_1 - \sigma_2}{\rho_1} = 0$$

the χ' is:

$$\chi' = \frac{1}{\sigma_1 - \sigma_2} \frac{d(\sigma_1 - \sigma_2)}{ds_2} - \frac{1}{\rho_1} \tag{7.18}$$

where ρ_1 is the curvature radius at the root of the notch.

Fig. 7.17 Direct measurement of χ' from photoelastic isochromatic fringes



The χ' is derived directly as inverse of the intercept \overline{AB} of the straight line on the s_2 axis, minus the inverse of ρ_1 :

$$\frac{\sigma_1 - \sigma_2}{\overline{AB}} = \frac{d(\sigma_1 - \sigma_2)}{ds_2} \quad \frac{1}{\overline{AB}} = \frac{1}{\sigma_1 - \sigma_2} \frac{d(\sigma_1 - \sigma_2)}{ds_2} \quad (7.19)$$

then:

$$\chi' = \frac{1}{\overline{AB}} - \frac{1}{\rho_1} \quad (7.20)$$

Exercise 7.1 (*Relative Stress Gradient for a circular hole in a thin plate; theoretical calculation and experimental determination*)

(a) Before solving the problem experimentally, let us develop the theoretical solution for a hole in an infinite plate loaded in one direction.

The principal stress $\sigma_1 = \sigma_y$ and the difference between the principal stresses $\sigma_y - \sigma_x$ along the loading axis, Eq. 6.38 (together with Eq. 6.33 of Chap. 6) are used for determining the χ' factor.

Theoretically, it is possible to calculate it in two ways: either through the two terms of Eq. 7.18 or directly through the single term of Eq. 7.17. In the first case $1/\overline{AB}$ is obtained as a limit of the ratio $\frac{1}{\sigma_1} / d(\sigma_y - \sigma_x) ds_2$ for x tending to a ,⁶ while in the second case χ' is directly obtained as a limit for x tending to a of Eq. 7.17.

Figure 7.18 shows plots of both derivatives that achieve their limit values for x tending to a , respectively equal to $+0.1852$ for the first case, and $+0.1296$ for the second case. Consequently, the second term $\frac{1}{\rho_1}$ of Eq. 7.18 must be 0.0556 .

(b) Given the image of Fig. 7.18b, determine the RSG according to Eq. 7.20.

From Fig. 7.17, the first term of the equation is $1/\overline{AB}$. In the present example it is 0.1852 $1/\text{mm}$, i.e. $\overline{AB} = 5.4$ mm. Then the theoretical ratio between the radius of the hole AC and AB is: $18/5.4 = 3.333$. Vice-versa, the ratio determined experimentally is 2.65 . This means that the value of \overline{AB} is $18/2.65 = 6.792$ and $1/\overline{AB} = 0.147$. This value is considerably lower than the theoretical one (0.1852), obtained for a large plate ($2a/H = 0$).

The experimental value of χ' , according to Eq. 7.20, is:

$$0.147 - 1/18 = 0.0914 \text{ mm}^{-1}$$

valid for $2a/H = 36/122 = 0.295$.⁷

Another approximate solution is given in Table 7.16. The χ' for a hole in a plane strip is given by the formula $2/\rho$ that in the present example gives $2/18 = 0.111$.

⁶ σ_x is zero for $x = a$.

⁷ Incidentally, Eq. 6.42 gives 3.35 , as maximum value of σ_y for $2a/H = 0.295$, a value about 10 % higher than the theoretical value for a large plate (3).

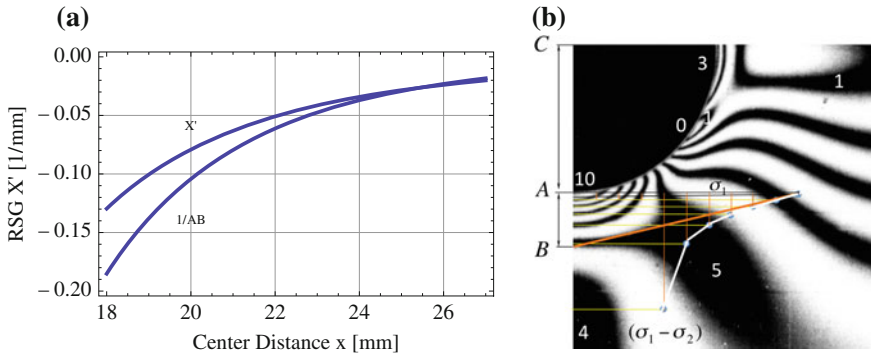


Fig. 7.18 a Plot of Eq. 7.17, (χ') and for the first term of Eq. 7.18, ($1/AB$) and b direct measurement of the χ' from photoelastic isochromatic fringes (External forces along the horizontal axis)

7.7.2 Relative Stress Gradient for Three Dimensional Bodies

Numerical and experimental methods are the proper tools for determining χ' at reference points of complex structures, in which three principal stress values are originated in terms of amplitudes and mean values, $\sigma_{1a}, \sigma_{2a}, \sigma_{3a}$ and $\sigma_{1m}, \sigma_{2m}, \sigma_{3m}$. Generally, reference points are located at the surface, and χ' is evaluated along the perpendicular direction for the stress components lying on the surface.

7.8 Dependence of Fatigue Strength on the Support Factor

Fatigue resistance at a reference point is different for uniform and non-uniform stress distribution because of the different support effect offered by the surrounding material. For example, the analysis of data in Table 7.1 [4] for a variety of metals shows this property: bending fatigue limits, with a χ' greater than zero are always bigger than the corresponding ones for push/pull load, Table 7.2.

A drawback of predicting models based on RSG estimation is that, if the stress distribution is very steep, it could be critical to estimate the gradient values. It would require the adoption of a finite-element model with very high mesh refinement with relatively long computational times, or an alternative very accurate high cost experimental analysis. Nevertheless, it must be observed that the support factor does not increase indefinitely as the gradient increases: the risk of an overestimation of the support factor is rare for high values of χ' .

Table 7.1 Fatigue limits versus static limits, from [4]

1 M	2 P/P ^a	3 P/P	4 Ben. ^a	5 Ben.	6 Ben.	7 Tor. ^a	8 Tor.	9 Tor.
	σ_{zdW}	σ_{zdo}	σ_{bW}	σ_{bo}	σ_{bs}	τ_W	τ_o	τ_s
1	$0.45R_m$	$1.3\sigma_W$	$0.49R_m$	$1.5\sigma_{bW}$	$1.5R_s$	$0.35R_m$	$1.1\tau_W$	$0.7R_s$
2	$0.41R_m$	$1.7\sigma_W$	$0.44R_m$	$1.7\sigma_{bW}$	$1.4R_s$	$0.30R_m$	$1.6\tau_W$	$0.7R_s$
3	$0.40R_m$	$1.6\sigma_W$	$0.41R_m$	$1.7\sigma_{bW}$	$1.4R_s$	$0.30R_m$	$1.4\tau_W$	$0.7R_s$
4	$0.25R_m$	$1.6\sigma_W$	$0.37R_m$	$1.8\sigma_{bW}$		$0.36R_m$	$1.6\tau_W$	
5	$0.30R_m$		$0.40R_m$			$0.25R_m$		

^a P/P = Push-Pull (Uniaxial Tension/Compression) Test

Ben. = Rotating Bending Test

Tor. = Alternate Torsional Test

1. Construction Steels
2. Hardened Steels
3. Case hardened Steels
4. Gray Cast Iron
5. Light Metal

Symbols

R_m = Ultimate Stress in Tension

R_s = Yielding Stress in Tension for 0.2 %

R_{bs} = Yielding Stress in Bending

τ_s = Yielding Shear Stress in Torsion

σ_{zdW} = Fatigue Amplitude Limit for Alternate Load ($R = -1$) in Push/Pull Test (The German symbols are adopted)

σ_{zdo} = Upper Fatigue Limit for ($R = 0$) in Push/Pull Test

σ_{bW} = Fatigue Amplitude Limit for Alternate Load ($R = -1$) in Bending Tests (with Moment in one plane or with rotating Moment)

σ_{bo} = Upper Fatigue Limit for ($R = 0$) in Bending Tests (with Moment in one plane)

τ_W = Fatigue Amplitude Limit with Torsional Alternating Moment ($R = -1$)

τ_o = Upper Fatigue Limit for ($R = 0$) in Torsion Test

a = Amplitude

zd = Tension/compression stress

b = Bending stress

W = Fully reversed stress ($R = -1$)

o = Stress at origin ($R = 0$)

s = Yielding

7.8.1 Size Effect

The relative stress gradient RSG, whose dimension is the inverse of a length, automatically takes into account the size effect: as the body dimensions increase, stress gradients and support factors decrease and larger volumes of material support higher stress levels. The most loaded fibers do not take advantage of local stress redistribution and the likelihood of presence of micro-structural defects also increases; both effects are responsible for a fatigue strength reduction.

Table 7.2 Comparison of fatigue limits for alternate bending and alternate push-pull load for columns 2, 4 of Table 7.1

No.	$\frac{\sigma_{bW}}{\sigma_{zdW}}$	Materials
1	1.09	Structural steels
2	1.07	Hardening and relieving
3	1.03	Case hardening steels
4	1.48	Grey cast iron
5	1.33	Light metals

7.8.2 Model of the Support Factor as a Function of the Relative Stress Gradient

Historically, the simulations were proposed by the researchers that first developed extensive experimental tests on this factor [34, 35], summarized in Fig. 7.19 together with the correspondent curves in log-log coordinates [22].

7.8.2.1 Siebel-FKM Model

The original model due to Siebel and Stieler [4, 34–36] is a function of two parameters, χ' (1/mm) and S_g . The last one is the size of the critical volume in a

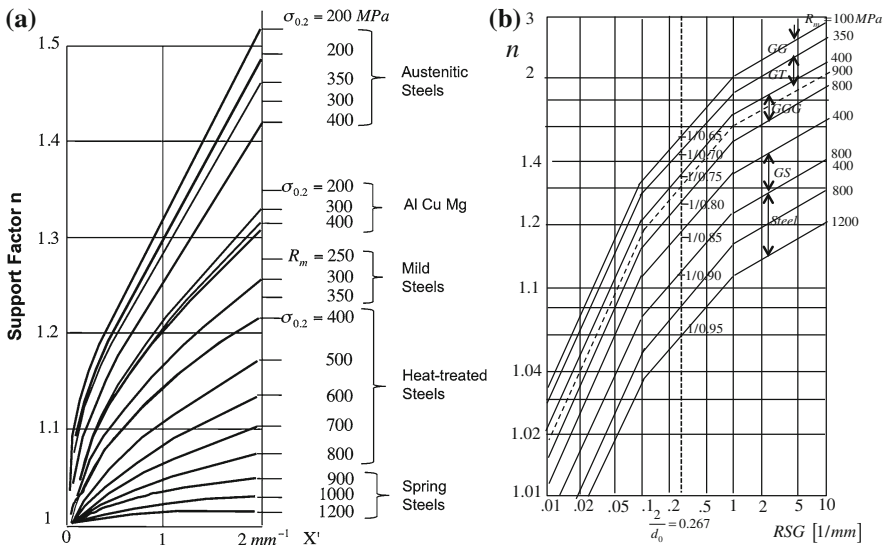


Fig. 7.19 Examples of curves of support factor (or dynamic stress reduction factor) for normal stress versus relative stress gradient for various metals on the left, **a** (after Siebel and Stieler) and on the right, **b** in logarithmic coordinates from [22]

perpendicular direction to the surface where the shear deformations are located and has the same order of magnitude as the grains diameter.

The following expression holds for the support factor:

$$n = 1 + \sqrt{S_g \cdot \chi'} \tag{7.21}$$

where S_g is equal to:

$$S_g = -0.0375 \cdot \ln R_m + 0.27 \tag{7.22}$$

(R_m in MPa).

The model suggested by FKM guidelines [22] distinguishes three fields of the support factor, for different relative stress gradient ranges:

For $\chi' \leq 0.1 \text{ mm}^{-1}$:

$$n = 1 + \chi' \cdot 10^{-\left(a_G - 0.5 + \frac{R_m}{b_G}\right)} \tag{7.23}$$

For $0.1 \text{ mm}^{-1} < \chi' \leq 1 \text{ mm}^{-1}$:

$$n = 1 + \sqrt{\chi'} \cdot 10^{-\left(a_G + \frac{R_m}{b_G}\right)} \tag{7.24}$$

For $1 \text{ mm}^{-1} < \chi' \leq 100 \text{ mm}^{-1}$:

$$n = 1 + \sqrt[4]{\chi'} \cdot 10^{-\left(a_G + \frac{R_m}{b_G}\right)} \tag{7.25}$$

Constants in the formula must be chosen as material functions, according to the Table 7.3.

- GS = Cast Steel,
- GG = Grey Cast Iron,
- GGG = Spheroidal Gray Cast Iron,
- GT = Hardened Cast Iron
- χ' in 1/mm and R_m in MPa.
- These formulas are valid at the reference points for a stress gradient normal to the surface, for normal stresses only. Numerical values in the diagram: 1/0.65 ÷

Table 7.3 a_G and b_G as a function of the kind of material

Mat.	Stainless steels	Other steel	GS	GGG	GT	GG
a_G	0.40	0.50	0.25	0.05	-0.05	-0.05
b_G	2400	2700	2000	3200	3200	3200

1/0.95 give the difference of the fatigue limit for reversed stress in tension-compression and in bending, for specimen of diameter $d_0 = 7.5$ mm, (and χ' equal to 0.267).

- In the case of slender bodies the diagrams and the models give the support factor that in this case is $n = K_t/K_f$. These factors are valid only for tension/compression.
- The same diagram is valid for a 3D-shaped body, where n is the ratio between the *local fatigue strength* for ($\chi' \neq 0$) and the *fatigue strength in a reference specimen* of the same material, in uniaxial loading condition ($\chi' = 0$), and also for massive details with 3D stresses, in slender bodies.
- The local values of component fatigue limit (for totally reversed stress) are obtained from the material fatigue limits, multiplying by the support factors.

Equation 7.24 is shown in Fig. 7.19b in double-logarithmic diagrams: (χ' in 1/mm and R_m in MPa).

For $\chi' \leq 0.1 \text{ mm}^{-1}$:

$$\log(n-1) = \log \chi' - \left(a_G - 0.5 + \frac{R_m}{b_G} \right) \quad (7.26)$$

For $0.1 \text{ mm}^{-1} < \chi' \leq 1 \text{ mm}^{-1}$:

$$\log(n-1) = 1/2 \cdot \log \chi' - \left(a_G + \frac{R_m}{b_G} \right) \quad (7.27)$$

For $1 \text{ mm}^{-1} < \chi' \leq 100 \text{ mm}^{-1}$:

$$\log(n-1) = 1/4 \cdot \log \chi' - \left(a_G + \frac{R_m}{b_G} \right) \quad (7.28)$$

From this point of view:

- The approach defines the concept of *component fatigue limits*; the fatigue limit is no longer a characteristic of a material but a property of a structural component that can have one or more reference points in which the fatigue life must be assessed.
- In 3D-components the calculation of the local stress state is supposed to be done by finite-elements methods or obtained by an experimental analysis, in order to evaluate principal stresses $\sigma_1 \sigma_2 \sigma_3$ at each reference point, in terms of amplitudes and means values.
- If the all stresses vary proportionally to the load and consequently to their amplitude and mean values, the directions of principal stresses remain unchanged and the principal stresses can be treated with the equivalence stresses.
- In the case of non-proportional stresses, it is necessary to distinguish the case of proportional amplitudes but non-proportional mean values from the more general case of total non-proportionality. In the first case, since the amplitudes are the dominant parameters of fatigue life, the problem can be re-conducted to the case of

proportional stresses. In the second case, approximate methods can be developed, but a direct experimental assessment is always recommended [22].

- At the reference point, that is generally located on the external surface of the component, if 1 and 2 are the principal stress directions on the surface and 3 is the direction normal to the surface, only gradients along 3 are taken into account for the calculation of the endurance fatigue limits σ_{a1} σ_{a2} . The diagram of Fig. 7.19a or b or the relative mathematical models, give the values of support factors for σ_{a1} and σ_{a2} , respectively $n_{\sigma_{a1}}$ and $n_{\sigma_{a2}}$.

For example, according to [22], the *fatigue limits* of 3D components, in directions 1, 2 and 3 are:

$$\begin{cases} \sigma_{a1,WDK} = n_{\sigma_{a1}} \cdot \sigma_{a1,zd DW} \\ \sigma_{a2,WDK} = n_{\sigma_{a2}} \cdot \sigma_{a2,zd DW} \\ \sigma_{a3,WDK} = \sigma_{a3,zd DW} \end{cases} \quad (7.29)$$

where index *a* means amplitude, the index *W* totally reverse stress and *K Component*, for which $\chi' \neq 0$.

7.8.2.2 Eichlseder Model

Unlike the [22] model, the simulation for the support factor in [27] is based on the knowledge of fatigue response of the same material under two reference fatigue tests, for χ' equal to zero (tension/compression of smooth specimen) and for $\chi' = 2/d_0$, (bending test). The $n = f(\chi')$ model [27] (recalled also in [37–42]) is the following:

$$n = \frac{\sigma_{aWDK}}{\sigma_{zd WD \chi'=0}} = 1 + \left(\frac{\sigma_{b WD}}{\sigma_{zd WD}} - 1 \right) \cdot \left(\frac{\chi'}{2/d} \right)^{K_D} \quad (7.30)$$

where:

- $\sigma_{aWD \chi' \neq 0} = \sigma_{aWDK}$ = local amplitude fatigue (endurance) limit at a reference point in the case of $\sigma_m = 0$.
- $\sigma_{zd WD \chi'=0}$ = fatigue (endurance) limit for the amplitude in a tension/compression test with $R = -1$ ($\chi' = 0$).
- $\sigma_{b WD}$ = fatigue (endurance) limit for the amplitude in a bending test with $R = -1$, ($\chi' = 2/d$)
- $\chi' = 2/d$ for the specimen in bending.
- K_D = Material coefficient for taking into account the non-linear behavior of the function, Fig. 7.20.

In double logarithmic variables Eq. 7.30 becomes a linear equation:

$$\log(n - 1) = \log\left(\frac{\sigma_{b WD}}{\sigma_{zd WD}} - 1\right) + K_D \cdot (\log \chi' - \log 2d) \quad (7.31)$$

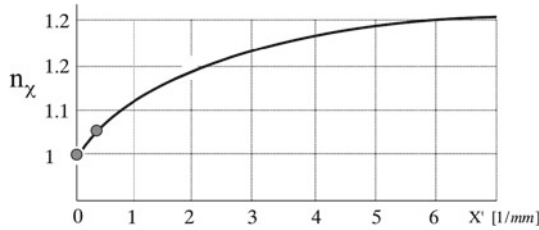


Fig. 7.20 Curve of support factor versus relative stress gradient derived from two reference tests of Push-Pull ($\chi' = 0$) and fully reversed bending moment ($\chi' = 0.267$ 1/mm)

7.8.2.3 Slope Modeling of the Local Fatigue Curve

From an idea of [27], modified in [43], a model has been proposed for simulating the S/N local curve. The local k value is inferred by elementary values of Push/Pull and Bending tests slopes, with an analogy to Eq. 7.30. If k_{zdW} is the slope of the fatigue curve obtained by a Push/Pull test, and k_{bW} is the slope of the fatigue curve obtained by a rotating bending test, the local value of k can be derived as, Fig. 7.21:

$$\frac{k_{locale}}{k_{zdW}} = 1 + \left(\frac{k_{bW}}{k_{zdW}} - 1 \right) \cdot \left(\frac{\chi'}{2/d} \right)^m$$

The value of m , for Low Alloy Steels is equal to 0.05.

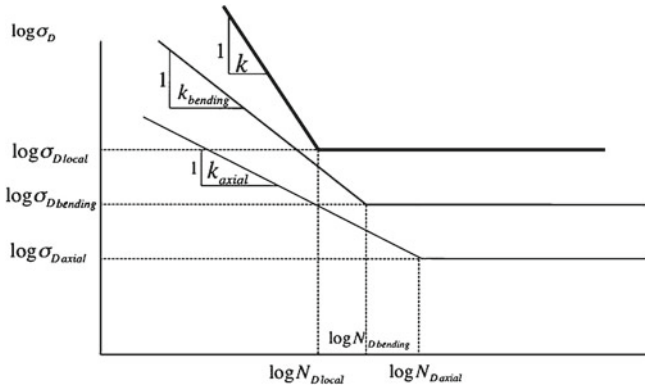


Fig. 7.21 Local fatigue curves at different χ' values, modified by [43] from [27]

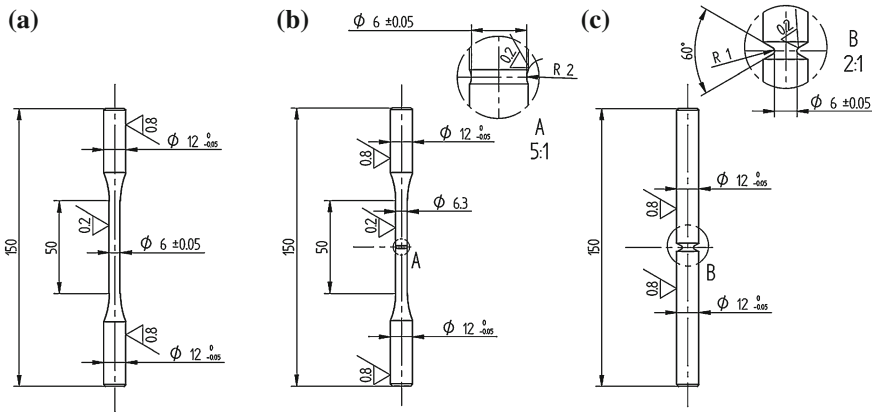


Fig. 7.22 Specimens geometry: smooth (a), shallow notched (b) and sharply notched (c)

7.8.3 Modeling Support Effect Through Rotating Bending Fatigue Tests: Case Study on Characterization of a Titanium Alloy

One difficulty in the FKM model is represented by the high number of specimens necessary to describe the whole useful field of relative stress gradient, together with the impractical use of two different testing machines.

As regards the simpler Eichlseder’s method, an additional observation is that the interpolation is performed between two χ' values that are close to one another.⁸ These values can be very different from χ' values corresponding to most notch geometries.

Establishing a unified theory mixing axial tests and bending tests, performed on specimens of different shapes and dimensions with different testing machines (push/pull and bending tests), is a weak point of the classic procedure.

A method is suggested utilizing only rotating bending tests for modeling the support factor of new materials [44]. The case study refers to the fatigue characterization of a material Ti-6Al-4V alloy, widely used in aerospace industries for its high specific strength, excellent fatigue properties and corrosion resistance, but not classified in fatigue handbooks.

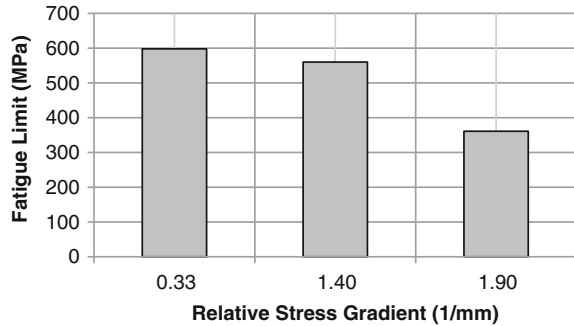
The Ti-6Al-4V grade 5 has an ultimate strength of 950 MPa and a yield strength of 880 MPa. Its composition is shown in Table 7.4.

Ti-6Al-4V specimens, with notches of different severity, were selected to demonstrate the assumptions, Fig. 7.22. The three kinds of specimens have a common reference diameter at the notch roots of 6 mm. The notched specimens were chosen

⁸ χ' is equal to 0 in tension/compression and 2 divided by the sample diameter in bending, which, considering most widely used diameters of about 6 ÷ 8 mm, gives a range of χ' from 0.33 to 0.25 1/mm.

Table 7.4 Chemical composition of the Ti6Al4V alloy

Element	Ti	Al	V	Fe	O	C	N	H
Content (%)	Balance	5.50–6.75	3.50–4.50	≤0.40	≤0.20	≤0.080	≤0.030	≤0.015

Fig. 7.23 Experimentally determined nominal fatigue limits for smooth, shallow and sharply notched samples**Table 7.5** K_f , K_t , n for three values of χ'

σ_{bDk}	σ_D	K_f	K_t	K_t/K_f	χ'	σ_{loc}
598	598	1.00	1.00	1.00	0.33	598
560	598	1.07	1.3	1.22	1.40	728
361	598	1.66	1.9	1.15	1.90	686

with known values of K_t and χ' [4, 45]. For $K_t = 1.3$ the relative stress gradient is $\chi' = 1.40 \text{ mm}^{-1}$ (shallow notched specimen) and for $K_t = 1.9$ is $\chi' = 1.90 \text{ mm}^{-1}$ (sharply notched specimen). For smooth specimens χ' is 0.33 mm^{-1} (Fig. 7.23).⁹

7.8.3.1 Predictive Model Based on Four Points

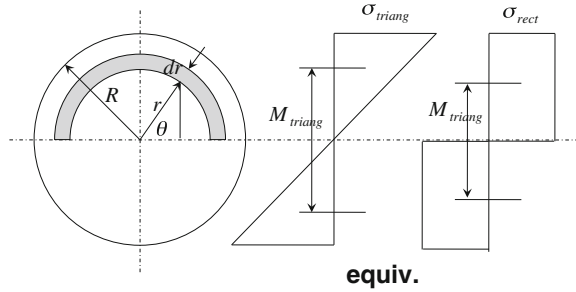
Table 7.5 shows the local values of the fatigue limit obtained multiplying the (nominal) fatigue limit σ_{bDk} of notched specimens by K_t .

In order to add the reference point for $\chi' = 0$ to this interpolation on three points (all obtained for $\chi' \neq 0$), the following escamotage is suggested:

- A fourth case is introduced of a virtual bending moment applied to a smooth specimen with χ' equal to 0. This case is equivalent to the tension/compression case with uniform stress distribution on the cross-section, Fig. 7.24.

⁹ The Dixon staircase method, Sect. 7.13 with $8 \div 10$ specimens for each specimen type and a stress step of 26 MPa is adopted, corresponding to about 5% of the expected fatigue limit, according to the literature suggestions.

Fig. 7.24 Equivalence between a bending triangular stress distribution and a bending uniform stress distribution on the same circular cross-section of the smooth specimen
 $(c = 16/(3 * \pi) = 16/(3 * 3.14) = 1.698$



The corresponding uniform stress value that is equivalent to the bending stress distribution on the same circular section is calculated. This is the new reference point for $\chi' = 0$. For the equivalence between the two distributions it is necessary to equalize the *static moments* with respect to the neutral axis of a triangular distribution and of a constant stress distribution, Fig. 7.24.

For a rectangular distribution the static moment with respect to the diameter is

$$M = \int_A \sigma_{rect} y dA$$

For a triangular distribution:

$$M = \int_A \frac{\sigma_{triangular}}{R} y \cdot y dA$$

where $dA = r dr d\theta$ and $y = r \sin \theta$.

Substituting and equalizing the two expressions:

$$\int_0^\pi \int_0^R \sigma_{rect} r \sin \theta r dr d\theta = \int_0^\pi \int_0^R \frac{\sigma_{triangular}}{R} \cdot r^2 \cdot \sin^2 \theta r dr d\theta.$$

From this equality we have:

$$\sigma_{rect} \cdot \frac{R^3}{3} \cdot 2 = \frac{\sigma_{triangular}}{R} \cdot \frac{R^4}{4} \cdot \frac{\pi}{2}$$

from which:

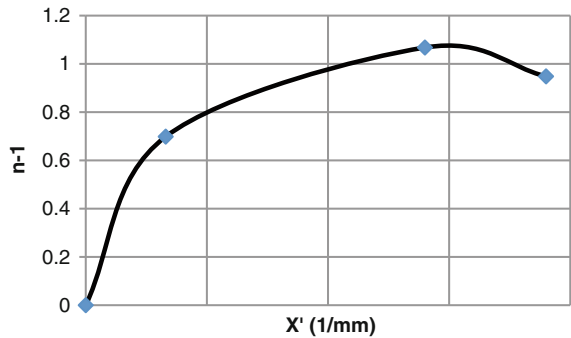
$$\sigma_{rect} = \frac{3}{16} \pi \cdot \sigma_{triangular} = \frac{\sigma_{triangular}}{c}$$

with the constant c equal to: $c = 16/(3\pi) = 1.698$

Table 7.6 Support factor $n = \sigma_{\chi' \neq 0} / \sigma_{\chi' = 0}$ as function of four values of RSG

σ_{bDk}	new K_t	n	χ'	$n - 1$
352	1.00	1.00	0.00	0.00
598	1.70	1.70	0.33	0.70
728	2.21	2.07	1.40	1.07
686	3.23	1.95	1.90	0.95

Fig. 7.25 Interpolating the spline of $(n - 1)$ versus χ' of four points



- The previous values of the bending stress found for the three geometries must be now referred to this new stress parameter, the virtual case of uniform stress value (e.g., the previous value of 598 MPa, is thus equal to $598/c = 352$); adding the new value for $\chi' = 0$ and $n = 1$ to Table 7.5 gives a new list of four local fatigue limits, shown in Table 7.6.

The interpolation of these new four values gives the mathematical models of the support factor on the whole domain for which the function n is defined. It is easy to see the difference between the spline that crosses the points, Fig. 7.25 and the fitting function in Fig. 7.26, $(n - 1) = C \cdot \chi'^m$ where C and m are the constants to be determined minimizing the functional of discrepancies between model and data.

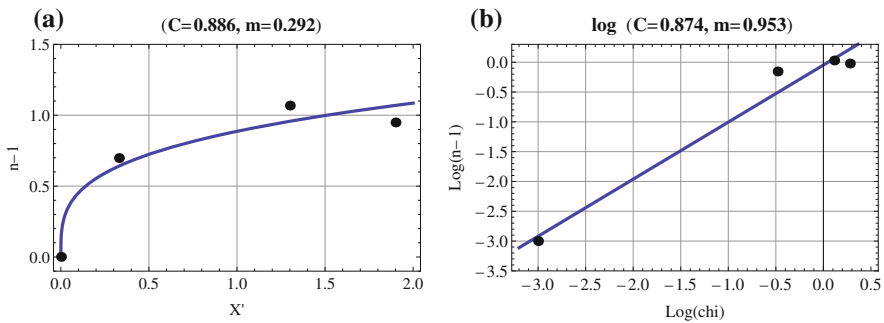


Fig. 7.26 **a** Fitting to four points with a non-linear function. **b** Fitting with the use of logarithms and a linear function

The minimum squares method gives the following result:

$$C = 0.886 \quad m = 0.292 \quad (7.32)$$

A linear fitting can be done for the logarithms of the data by the equation:

$$\log(n - 1) = \log C + m \log x \quad (7.33)$$

A fitting obtained by logarithms equation offers approximate values of the parameters, Fig. 7.26b as will be shown later:

$$C = 0.874 \quad m = 0.953 \quad (7.34)$$

7.8.3.2 Shot-Peening Treatment for Finding Other n Values

A curve relating n to the RSG can be extended to consider the shot-peening effect as a source of increment of the RSG.

$$\chi' = \frac{1}{(\sigma_x - \sigma_{res.})} \frac{d\sigma_x}{dx} \quad (7.35)$$

Residual stress distribution can be carefully estimated by diffractometric measurements. The in-depth measurements required step-by-step removal of thin material layers: this operation was performed by using an electropolishing facility to prevent considerable alteration of the pre-existent residual stress state; details in [44].

Fatigue tests were performed on two specimen types: respectively smooth and sharply notched. Three different treatments were experimented on them:

1. *Almen* intensity of 12N, with ceramic shots having a diameter of approximately 150 mm with a 200 % coverage (Z150-12N).
2. *Almen* intensity of 6A, with steel shots having a diameter of approximately 280 mm with a 200 % coverage (S110-6A).
3. *Almen* intensity of 12A, with steel shots having a diameter of approximately 580 mm with a 200 % coverage (S230-12A).

The results of the fatigue tests are shown in the histogram in Fig. 7.27. The data are shown in Table 7.7. Also in this case it is necessary to refer all the local maximum stresses to the virtual case of uniform stress value and the new Table 7.8 is determined.

The interpolation curves are shown in Fig. 7.28. It is enough to add only one point, (for a total of five points), obtained by shot-peening of a smooth specimen with $\chi' = 0.33 \text{ mm}^{-1}$ to considerably improve the model that fits well to the all ten experimental data. The most reliable mathematical model is the following:

$$C = 0.849 \quad m = 0.239 \quad (7.36)$$

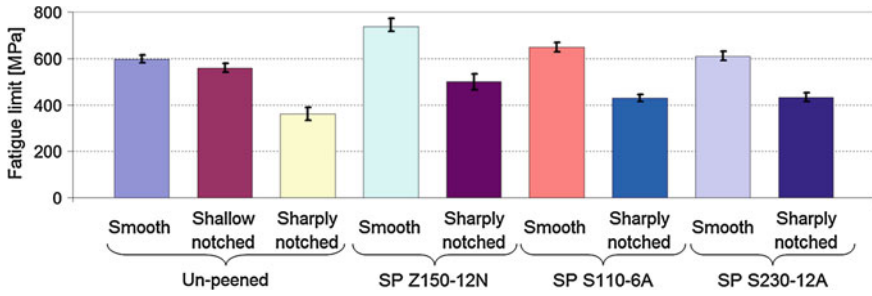


Fig. 7.27 Nominal fatigue limits for shot-peened smooth, shallow and sharply notched samples

Table 7.7 Data for n and χ'_{tot} calculation

Notch	Shot-pee	χ'_{load}	K_t	σ_D	σ_{Dk}	$(d\sigma/dx)_{load}$	$\sigma_{res.}$	σ_{tot}	χ'_{tot} load + res.str.
No notch	no	0.33	1	598	598	199	0	598	0.33
Mild	no	1.40	1.30	560	756	1008	0	756	1.33
Severe	no	1.90	1.90	361	635	1207	0	635	1.9
No notch	Z150-12N	0.33	1	736	736	254	-695	41	5.97
Severe	Z150-12N	1.90	1.90	500	879	1671	-695	184	9.07
No notch	S110-6A	0.33	1	648	648	254	-528	120	1.80
Severe	S110-6A	1.90	1.90	428	754	1433	-527	226.2	6.33
No notch	S230-12A	0.33	1	610	610	204	-532	79	2.58
Severe	S230-12A	1.90	1.90	433	762	1448	-531	-531	6.28

Table 7.8 Ten points Data

χ'	σ_{Dk}	n	n - 1
0	352.3	1	0
0.333	598.0	1.698	0.698
1.400	728.0	2.067	1.067
1.900	685.9	1.947	0.947
5.976	775.2	2.201	1.201
9.075	879.5	2.497	1.497
1.803	719.4	2.042	1.042
6.334	793.7	2.253	1.253
2.581	709.8	2.015	1.015
6.283	828.3	2.352	1.352

Exercise 7.2 (System Parameter Identification for the Support Factor Determination) In order to clarify the general criticality of building a mathematical model through linearized logarithmic relationships, several minimization algorithms are used, with the dubious result of obtaining different minima values with different

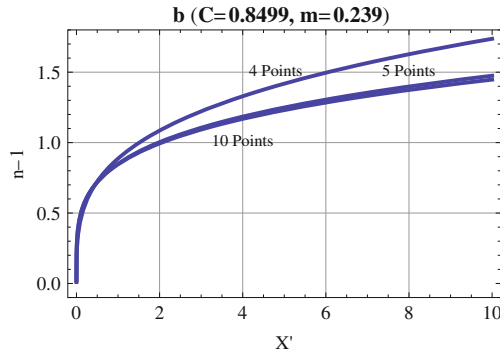


Fig. 7.28 Interpolating curve for $(n - 1)$ as a function of χ' in five points and comparison with four and ten points

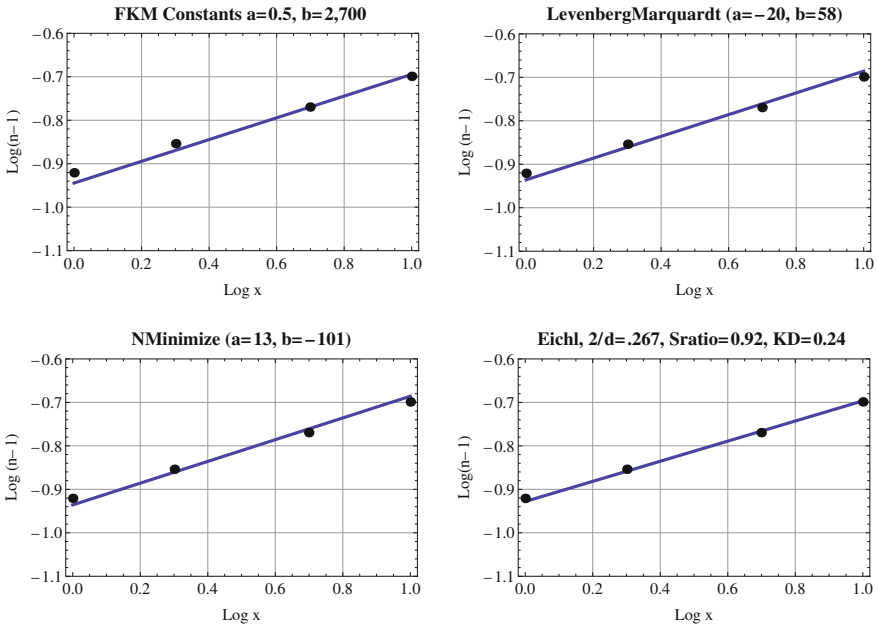


Fig. 7.29 Comparison between fitting models, varying algorithms of the optimization method

minimization techniques for finding factors a_G and b_G in the FKM model and also the previous K_D values. Fitting of straight lines to experimental data is shown in Fig. 7.29.¹⁰

¹⁰ FindFit of Mathematica® by default finds a least-squares fit but possible settings are foreseen for other methods, including “ConjugateGradient”, “Gradient”, “LevenbergMarquardt”, “NMinimize”, and “QuasiNewton”, being the default “Mean Square Method”.

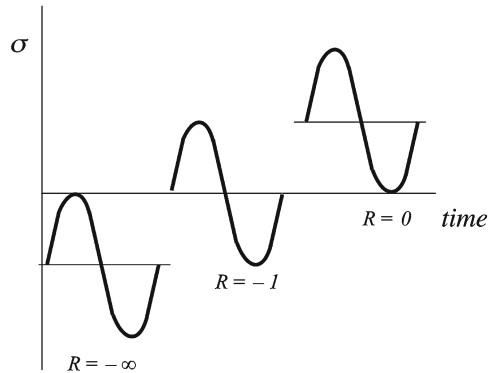


Fig. 7.30 Stress cycles at different R values; the sinusoidal shape is only indicative but not necessary for the theory. The cycle can consist of segments of *straight line*. $R = 0$ pulsating tension $R = -1$ fully reversed tension $R = -\infty$ pulsating compression

7.9 Models for the Evaluation of Mean Stress Effect

Figure 7.30 shows the main reference tests with three kinds of mean stresses.¹¹ In mechanical components a steady mean stress superimposed on the cyclic loading is very common. The mean stress has a strong influence on fatigue behavior: as we will see, tensile mean stresses are detrimental and compressive mean stresses are beneficial.

In order to describe this behavior in a mathematical way, a synthesis of a large number of experiments was done to build acceptable models for phenomenon simulation. The main points of this process are the following:

- The first experimental evidence of the influence on fatigue of the mean stress is shown in Fig. 7.31, where stress amplitude σ_a is plotted versus number of cycles to failure for different mean stress values σ_m . A positive value reduces fatigue limit while a negative value raises it. The central curve is the relationship between the stress amplitude for full inversion (mean stress equal zero) and cycle number to failure obtained by tests on smooth specimens, loaded in tension/compression, bending or torsion modes are always the fundamental information on the fatigue behavior, since the fatigue strength is determined in the first place by the stress amplitude.
- Investigations performed comparing the fatigue limits σ_{aN} versus the cycles numbers N with the mean stress, σ_m , for tension/compression and τ_m for torsion, give

¹¹ $\sigma_m = \frac{\sigma_o + \sigma_u}{2}$ mean stress
 $\sigma_a = \frac{\sigma_o - \sigma_u}{2}$ amplitude
 $\Delta\sigma = \sigma_o - \sigma_u$ range
 $f = 1/T$ frequency
 $R = \sigma_u/\sigma_o$ stress ratio.

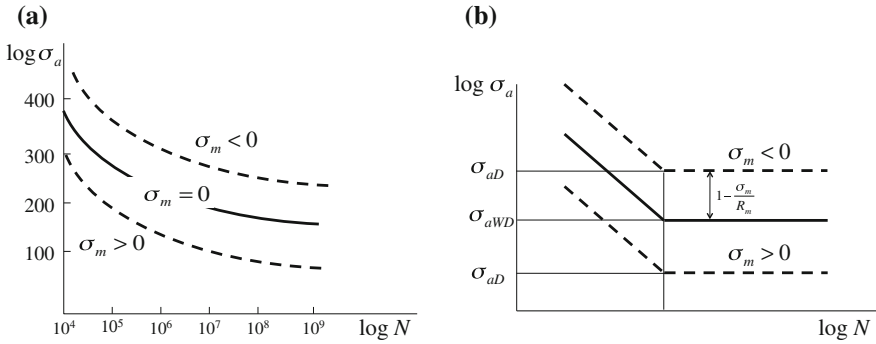


Fig. 7.31 Effect of a mean stress on endurance limit shown in semi-log (a) and log-log (b) diagrams

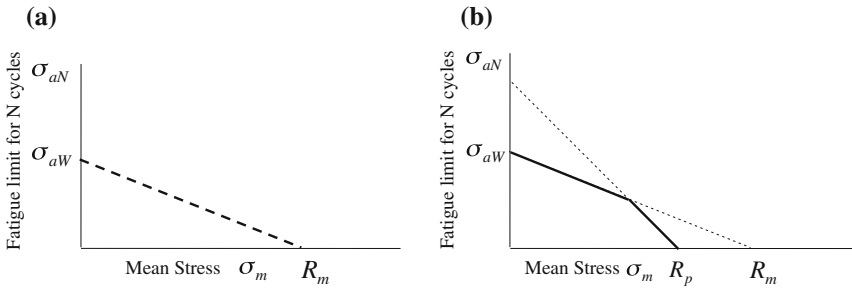


Fig. 7.32 First approximation (a) and second approximation (b) of Haig plots on the influence of mean stress on fatigue life for N cycles

the experimental results shown in Fig. 7.32a, b.¹² The plots, called Haigh diagrams, in Fig. 7.32a are described by the following equation:

$$\frac{\sigma_{aN}}{\sigma_{aW}} + \frac{\sigma_m}{R_m} = 1 \tag{7.37}$$

from which:

$$\frac{\sigma_{aN}}{\sigma_{aW}} = 1 - \frac{\sigma_m}{R_m} \tag{7.38}$$

The ratio between $\sigma_{m \neq 0}$ and endurance limit for alternate stress ($\sigma_{m=0}$) is:

$$\frac{\sigma_{aD}}{\sigma_{aWD}} = 1 - \frac{\sigma_m}{R_m} \tag{7.39}$$

¹² The models are valid for the endurance limit, for finite life as well as for long life when the endurance limit does not exist. In the latter case, σ_{aN} is the fatigue limit for a fixed long life N .

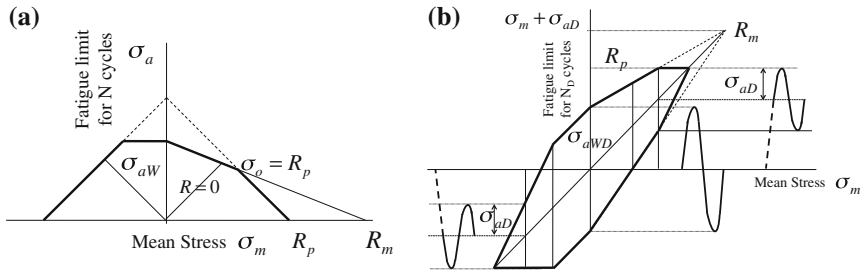


Fig. 7.33 Haig (a) and Smith (b) plots for several values of the mean stress

This is an approximate model, modified by several researchers, to improve the fitting to experimental data that this model tends to underestimate.¹³

An improvement consists in limiting the plot to the yield point $R_{p0.2}$, Fig. 7.32b, the plastic behavior being excluded by the elastic field of High Cycle Fatigue.

- According to the previous observations, the simplest (even approximate) way to take into account the influence of the mean stress on the local fatigue limit of Eq. 7.29 is to modify them in the following way, Eq. 7.38:

$$\begin{cases} \sigma_{a1,DK} = n_{\sigma_{a1}} \cdot \sigma_{a1zdDW} \cdot \left(1 - \frac{\sigma_m}{R_m}\right) \\ \sigma_{a2,DK} = n_{\sigma_{a2}} \cdot \sigma_{a2zdDW} \cdot \left(1 - \frac{\sigma_m}{R_m}\right) \end{cases} \quad (7.40)$$

- Another illustration of the modification of the endurance fatigue limit when a mean tensile or compressive stress is over-imposed is the Smith-Goodman diagram, Fig. 7.33b. It plots the sum of the mean stress σ_m and the stress amplitude $\sigma_W D$ versus the mean stress, up to the yield point $R_{p0.2}$.

A negative low mean stress favors a closure of the micro-cracks. Conservatively, as a first approximation, Eq. 7.38 can be used to model the fatigue curve, in this way:

$$N \cdot \sigma_a^k = N_D \cdot \sigma_{aD}^k = N_D \cdot \left(1 - \frac{\sigma_m}{R_m}\right)^k \sigma_{aWD}^k = A_1 \quad (7.41)$$

Comparing this equation with the Eq. 7.2 an elementary model of mean stress effect can be represented in double logarithmic plot as shown in Fig. 7.32a.

Equation 7.41 gives the elementary approximated influence of the mean stress on the fatigue curve. It can be written, N being equal to $= N_D$ for all the fatigue curves with different value of σ_m , as:

$$\sigma_a = \sigma_{aWD} \cdot \left(\frac{N_D}{N}\right)^{-1/k} \cdot \left(1 - \frac{\sigma_m}{R_m}\right) = \sigma_{aWD} \cdot \left(1 - \frac{\sigma_m}{R_m}\right) \quad (7.42)$$

¹³ A model suggests, e.g. the true fracture strength σ'_f [5] instead of R_m .

or:

$$\sigma_a = \sigma_{aWD} \cdot K_{AK} \quad (7.43)$$

with:

$$K_{AK} = \left(1 - \frac{\sigma_m}{R_m}\right) \quad (7.44)$$

The FKM guidelines [22] give a more detailed Haigh diagram for steels that differentiates the mean stress influence as a function of the way in which the component reaches the failure condition (at constant mean stress or at constant stress ratio or at constant minimum or maximum stress) as a function of the stress ratio R . Summarizing, the design against (high cycle) fatigue requires:

1. Calculation of the amplitudes of stresses (or equivalent ones) at the reference points.
2. Simulation of Local Fatigue Limit, i.e. the evaluation of the local strength at critical points of a mechanical component

The simulation proceeds according to the following steps:

- Determination of the endurance fatigue limit of the base material through laboratory tests on classic tension/compression as well as on bending specimens.
- Evaluation of the relative stress gradient in the reference points (inspected points) of the component. Since this parameter is defined in the elastic field, finite elements methods are suited to this task, even if the numerical estimation of a gradient, i.e. of a derivative of a function defined on a discrete domain, can present uncertainties.
- Evaluation of the support factors and the relative local fatigue limits at the reference points.
- Mean stress correction on the fatigue limit (amplitude).
- Comparison of these values with the principal stresses at those points.

7.10 Fatigue Assessment of Mechanical Components Through Relative Stress Gradient Theory

The following case study is an example for evaluating the fatigue life through the theory of Relative Stress Gradient. The results of the two theories will be presented.



Fig. 7.34 Three basic configurations: low, ground level and high

7.10.1 Case Study for the Fatigue Assessment at a Three-Dimensional Protrusion in an Excavator Arm

After an analysis of different working geometries of the excavator arm, the load configurations and the reference points in the component are selected for the maximum loading condition, Fig. 7.34:

1. Excavator at low position
2. Excavator at ground level
3. Excavator at high position

The reference point is localized close to a protrusion at which a hydraulic cylinder is connected. The theory for a shell-shaped (2D) component is not suitable for this case, while the local stress approach in a 3D area has to be considered preferable.

7.10.1.1 Calculation of the Principal Stresses at a Reference Point

The determination of the equivalent stress according to *Von Mises* theory (and of the max. principal positive stress), are performed in the most critical geometrical discontinuities, (called reference points), by the Finite Element Method for three loading conditions, see details in [46, 47]. Figure 7.35 shows the map and the Table 7.9 the relative values, of the equivalent stress for two of the three loading conditions.

7.10.1.2 Determination of the Local Strength at Each Reference Point

The strength at the reference points is determined by the application of the previous models according to the following steps:

1. Determination of the relative stress gradient at the reference points.
2. Evaluation of the support factors and of the local fatigue limits at the reference points.

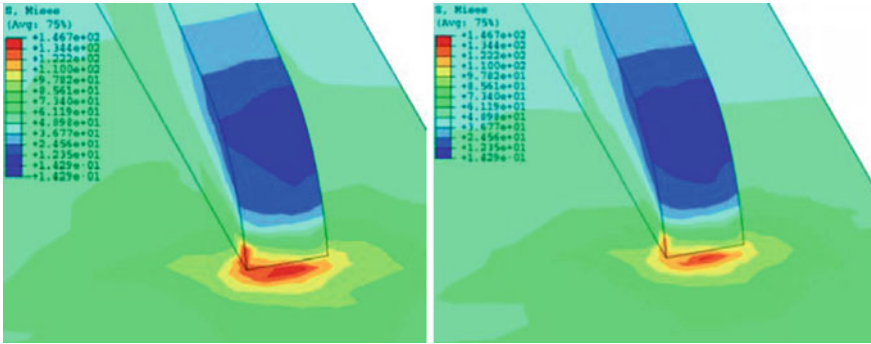


Fig. 7.35 Maximum equivalent stress (Von Mises) in the excavator arm

Table 7.9 Comparison of stress levels at the most critical point for the three loading conditions of the excavator arm

Condition	σ_{1max} (MPa)	σ_{1mean} (MPa)	σ_{1min} (MPa)	Von Mises (MPa)	Excavator position
1	166	42	4	147	Excavating at low level
2	103	27	3	91	Excavating at ground Level
3	-4	-38	-154	137	Excavating at high level

Evaluation of the Relative Stress Gradient at the Reference Points

The stresses at two points along the thickness at a distance Δx from one another (the first point on the surface and the second point close to the surface) is done by the Finite Element Method and the stress gradient is calculated through a finite difference estimation. From the difference of their spatial coordinates, the relative equivalent stress gradient along the direction \bar{x} perpendicular to external surface is given by:

$$\chi' = \frac{1}{\sigma_{max}} \frac{\Delta\sigma}{\Delta\bar{x}}$$

with:

$$\Delta\sigma = \sigma_1 - \sigma_2$$

and:

$$\Delta\bar{x} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

where 1 is a node at the surface and 2 a node immediately under the surface; σ is the Von Mises equivalent stress; and x, y, z are the spatial coordinates of the nodes. Figure 7.36 shows the equivalent stress distribution close to the reference point and the distance calculation in the selected point, Table 7.10. The section is identified by taking as a first point the node on the surface and as the second the node on the lower

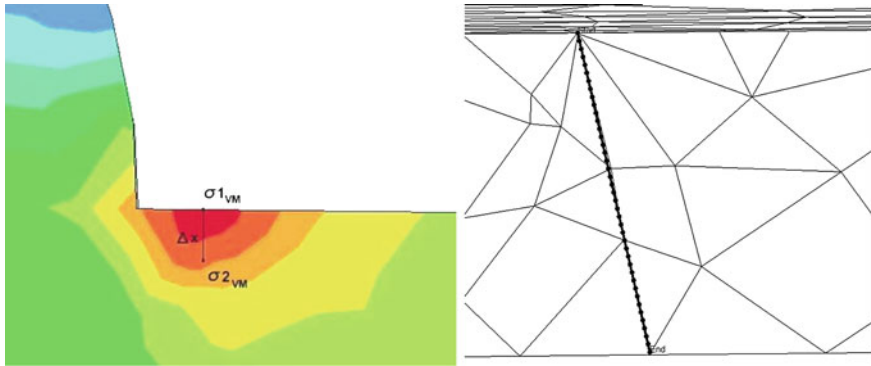


Fig. 7.36 Stress Gradient Calculation by FEM

Table 7.10 Spatial Coordinates of the nodes and Δx calculation (mm)

x_1 (mm)	x_2 (mm)	y_1 (mm)	y_2 (mm)	z_1 (mm)	z_2 (mm)	Δx (mm)
-54.89	-55.27	642.0	641.2	-1818	-1817	1.534

surface of the sheet that constitutes the thickness of the tubular section of the arm, choosing the closest node to an orthogonal line, Fig. 7.36b. The segment was divided into 40 intervals, and for each of them the FEM determined the principal stresses and the stress gradient for the Von Mises equivalent stress, Table 7.11.

Due to the automatic meshing algorithm, it could happen that the nodes of the elements under the surface do not rigorously stay on the line perpendicular to the surface. Nevertheless, due to the high number of the elements, the discrepancy between the correct values on the real perpendicular line and the computed values in the nearest nodes is very small.

Evaluation of the Support Factors and of the Local Fatigue Limits

The component is built in Steel *EN : S 355 (CNR – UNI : Fe 510 or DIN : St 52 – 3)*. The support factors can be estimated by several models as a function of χ' . The first model that is used in this example [39] is determined by Eq. 7.30. It is based on the knowledge of two material endurance fatigue limits for tension/compression and for bending moment, plus a third parameter K_D , typical for a class of materials.

Table 7.11 Example of a finite element calculation of principal stresses and of RSG of the maximum principal stress at the reference point, corresponding to three loading configurations

Configuration	σ_{1max} (MPa)	σ_2 (MPa)	σ_{max} (MPa)	$\Delta\sigma_{equiv.}$ (MPa)	$\chi = \frac{\Delta\sigma}{\sigma_{max}}$ (MPa/mm)	$\chi' = \frac{1}{\sigma_{max}} \frac{\Delta\sigma}{\Delta x}$ (1/mm)
1	147	133	147	14	9.13	0.0621
2	91	83	91	8	5.22	0.0573
3	137	125	137	12	7.82	0.0571

Table 7.12 Comparison of local strength at the reference points for three loading configurations of the excavator arm

Load	n_x	$\sigma_{aWD} \chi' \neq 0$	σ_{mean}	K_{aK}	$\sigma_{aDK} \chi' \neq 0 = \sigma_{aWD} \cdot K_{aK}$
1	1.201	269	42	0.93	250
2	1.196	268	27	0.95	255
3	1.196	268	-38	1.06	284

$\sigma_{aDK} \chi' \neq 0 = \sigma_{aWD} \cdot K_{aK}$ with $K_{aK} = \left(1 - \frac{\sigma_m}{R_m}\right)$, $R_M = 597$ MPa, $R_{p0.2} = 400$ MPa

The local fatigue endurance limit obtained by this model is shown in Table 7.12. Other models, such as Eq. 7.21 or FKM equation or the plots of Fig. 7.19, give other estimation values, shown in Table 7.13.

The model needs the preventive knowledge of the following basic fatigue values:

- Ratio $\sigma_b W / \sigma_{zd} W \chi' = 0 = 1.07 \div 1.09$, see Table 7.2 for ($\chi' = 2/d = 0.33$)
- $K_D = 0.3$ for this kind of steel and the corresponding n values in Table 7.13.

7.10.1.3 The Mean Stress Effect

The correction for the mean stress effect on fatigue life is done with the simplest model, Eq. 7.44. Table 7.12 shows the values of K_{aK} for different σ_m and the final estimations of the local strength (fatigue endurance limit for the component K), expressed as amplitude $\sigma_{aDK} \chi' \neq 0$ with its mean value.

7.11 Macroscopic Fractography as Stress Analysis Tool

In the second half of the 19th century, the description of the fracture surfaces became a useful integration in understanding fatigue behavior. Rankine made accurate observations on fractures of railway axles in Britain.¹⁴ This description still holds true

Table 7.13 Support Factor n estimated by Eq. 7.21 and diagram of Fig. 7.19

Equation 7.21 and Table in [4]	Figure 7.19 [22, 33]	Equation 7.30 [39]
1.05	$1.03 \div 1.05^a$	$1.04 \div 1.05^a$

^aMean value of three values of Table 7.11

¹⁴ William John Macquorn Rankine (1820–1872), Professor at the University of Glasgow, clarified the phenomenon of fatigue in fractured axles in railway vehicles: “...the broken end of the journal was convex, and necessarily the body of the axle was concave, until the thickness of sound iron in the centre became insufficient to support the shocks to which it was exposed. It is therefore proposed, in manufacturing axles, to form the journals with a large curve in the shoulder, before going to the lathe, so that the fibre shall be continuous throughout” [48].

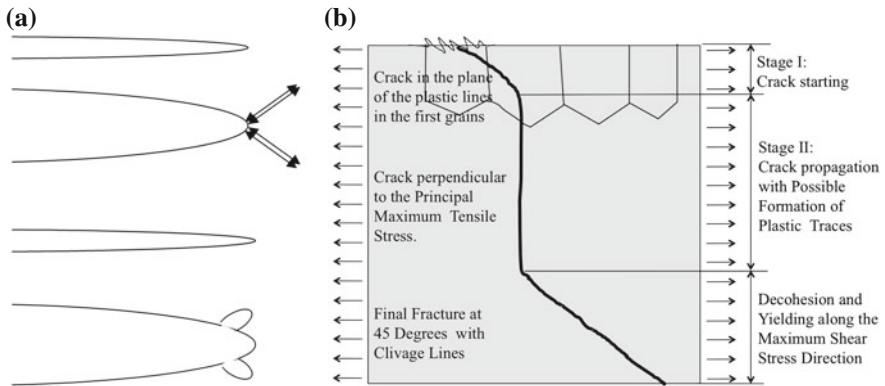


Fig. 7.37 From the *top left* Crack not loaded; crack loaded with the formation of flow lines at the tip at 45°; closing of the crack; advancement of the tip with residual traces of the first plastic shear; **b** different stages of crack propagation

as can be seen in a railway disaster where an axle broke in proximity of a wheel, Fig. 7.40.

Load cycles generally leave markings on the fractured surfaces, due to plastic deformation at the crack tip generated by the crack progress. This generates a succession of so called *beach marks*.¹⁵ The formation of *beach marks* can be explained by the model shown in Fig. 7.37 on the left. A formation mechanism of the marks on the fracture surface is shown, while, on the right, the two typical stages of propagation (stage I and stage II) plus a final stage of fracture are shown: at the beginning the process of fatigue is dominated by the maximum shear stress agent at 45° with respect to the load direction and after that by the maximum tensile stress. Close to the final rupture, the fracture follows a 45° direction, again controlled by shear stresses.

Macroscopic fractographic analysis describes the formation of marks limited to stage II, while in the first stage they do not occur and for some materials are not formed at all even in the second stage.

An example of this behavior is shown in Fig. 7.38. Qualitative but useful information can be deduced on the state of stress from the careful observation of the fracture surfaces and the shapes of the plastic beach marks:

- With the simple failure hypothesis of the maximum tensile stress (that is substantially verified for many structural materials), it is possible to conclude that the crack follows the shape of the second isostatics curve, materialized by a family of the isostatics curves seen in Chap. 3.
- The beach marks are the traces of the crack front during the propagation phase. They follow the laws of propagation examined in Chap. 9: the speed of a point

¹⁵ These marks, called also *arrest lines* are cancelled only if the fracture surfaces are cyclically in contact.



Fig. 7.38 Failure surface of a curved bar. The enucleation point, the propagation area and the final brittle failure area at 45° can be detected

at the front of the crack strongly depends on the Stress Intensity Factor in that direction.

- Let us consider a two-diameter shaft with an internal corner between the two parts. Fractures nucleate close to the corner. The mark curvature on the fracture surface depends on the stress level and on the three-axiality of the stress state, compare Fig. 9.7 of Chap. 9: at the internal points of the mark line, characterized by a plane strain state, the brittle propagation is favored, while at the surface points, in plain stress state, the formation of a plastic barrier is present, see Figs. 9.4 and 9.5. The shape of the beach mark is thus a balance of two different mechanisms: the surface bi-axiality with a high stress level due to the concentration factor in the corner and the internal three-axiality sustained by a lower principal stress level. According to the sharpness of the corner, one of the two effects can prevail, originating different curvatures for different cases.

7.11.1 Rotating Bending of a Shaft with Two Diameters

- In the case of a large corner radius between the two parts of the shaft, the maximum stress concentration is localized at the shaft surface with the smaller diameter, due to the stress trajectories shape that tend to be concentrated as shown at the top of Fig. 7.39. The appearance of the fracture surface, which is flat, see dotted line, is diversified for low and for high nominal stress, because in the first case marks arise from a single nucleation point while in the second case they have multiple origins. It is also possible to recognize the rotation direction of the shaft: in fact, the marks' asymmetry is due to the opening mode of the crack that is favored by the first quarter of the sinusoidal stress cycle that tends to open the crack in the phase of a rising stress.
- Conversely, in the case of a sharp corner, the maximum stress concentration occurs in the corner itself.

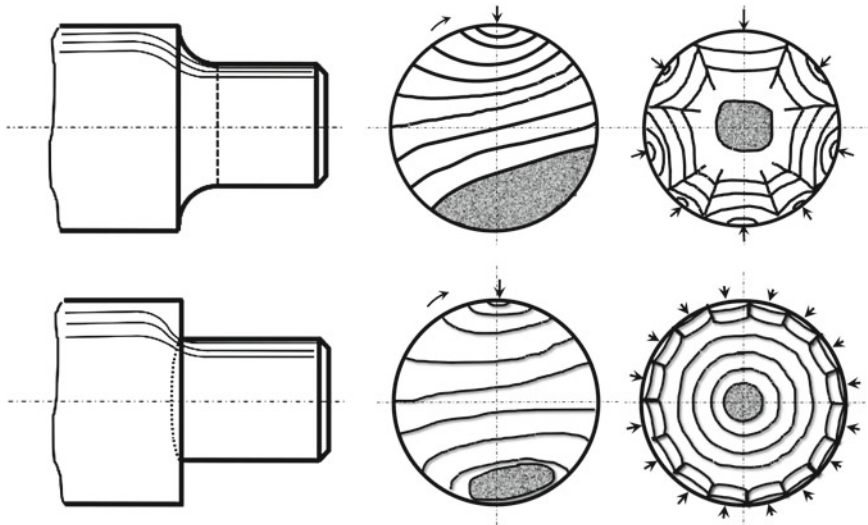


Fig. 7.39 Comparison between the position and the shape (*dotted lines*) of the fracture surfaces for rotating bending in the case of smooth (*top*) and sharp (*bottom*) corners, for low (*left*) and high (*right*) nominal stress



Fig. 7.40 Concave and convex surfaces in an axle in a railway failure

The appearance of the fracture surface, which is slightly curved and internal to the larger cross-section, Fig. 7.40, differs for low and high nominal stress, respectively as regards the origin of the progress lines from single or from several points. In the first case, propagation is faster at the surface where the stresses are highest with consequent change of the concavity of the marks. In the second case, the propagation tends to be uniform over the entire surface of the fracture.

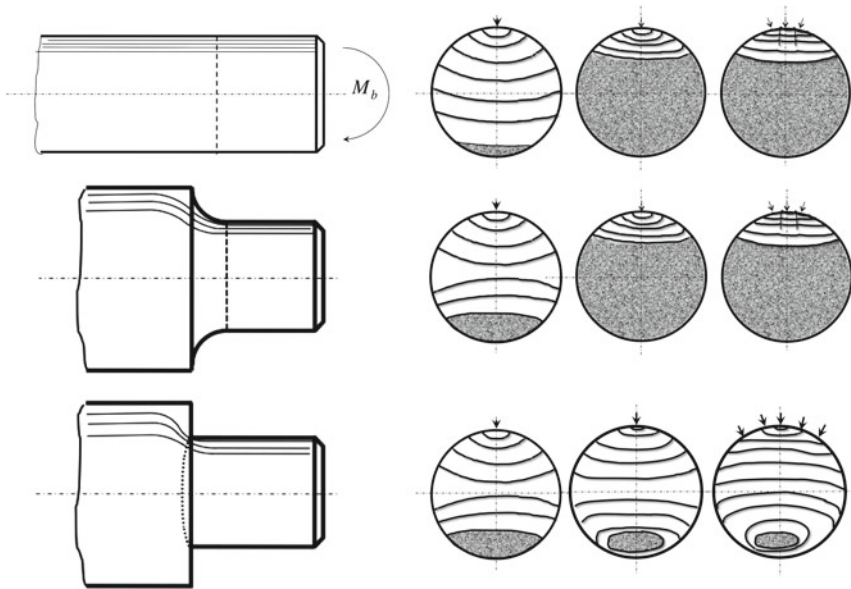


Fig. 7.41 Comparison between the position and the shape (*dotted lines*) of the fracture surfaces for unidirectional bending $R = 0$ in the case of uniform section (*top*), smooth (*middle*) and sharp (*bottom*) corners, for low (*left*) medium (*center*) and high (*right*) nominal stress

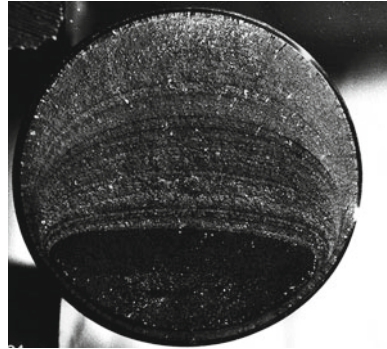
7.11.2 Shafts in Bending at a Stress Rate $R = 0$

If the shaft cross-section is uniform along the length, the maximum bending stress is constant along the surface. The trigger point is the weakest point, often due to a defect at the surface. The curved and symmetrical shape of the marks indicates the unique origin of the crack, and the concave shape of the beach mark that remains during the propagation is a sign of a propagation speed that is higher internally, characterized by states of three-axial stresses. The single origin and the final small residual area statically broken are typical of a propagation at low nominal stress, Fig. 7.41. A large final area conversely indicates a higher stress than the stress of the previous case. Further, this surface is often characterized by multiple triggers and radial grooves between the enucleation points and final closure of the marks. All of these points indicate a high stress, Fig. 7.42.

In the case of geometrical discontinuity with a low stress concentration, a bending moment applied in one direction ($R = 0$) gives rise to three cases:

1. the stress concentration in the surface fibers causes an increase in the propagation speed at the surface, for which the marks' concavity tends to change direction. The single-source and the reduced final area both indicate a propagation at low nominal stress.

Fig. 7.42 Example of fracture surface in a sharp corner for ($R = 0$) bending and high stress level



2. A final extended fractured area conversely indicates a nominal stress higher than the previous case.
3. The case of multiple enucleation points and grooves between the points and final connection of the marks indicates a very high nominal stress.

In the case of a sharp corner with high stress concentration, the crack propagates in the internal direction of the biggest portion of the shaft, as shown in Fig. 7.41, following the stress lines and the bending, gives rise to three cases, depending on its level: slow mark convexity changes due to the higher surface stress than internal stresses.

Faster convexity changes due to the higher surface stress, until marks surround the area of final fracture.

Faster convexity changes due to the even higher surface stress, until marks surround the area of the final fracture, with the nucleation starting from several surface points.

7.11.3 Shafts in Bending at a Stress Rate $R = -1$

The previous description can be extended to the case of a bending in the vertical plane Fig. 7.43, acting in both directions, $R = -1$. An example of a broken cylindrical pin with constant cross-section and constant bending plane, subjected to low nominal stress is shown in Fig. 7.43b.

7.11.4 Shafts in Torsion at Shear Stress Rate $R = -1$

Figure 7.44 shows two types of fracture for alternate torsion in both directions. The first figure on the left side is characteristic of the fracture with a high number of cycles

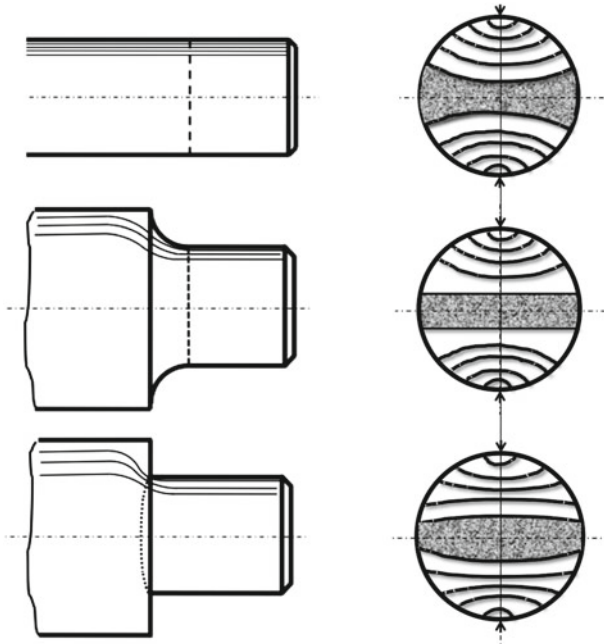


Fig. 7.43 Fracture surface for shafts in bending at stress rate $R = -1$

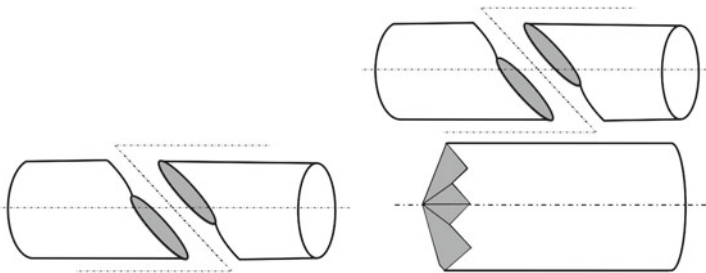


Fig. 7.44 Comparison between the fracture surfaces of a specimen with a circular cross-section loaded in torsion statically and dynamically, at high number of cycles

and low stress level, in which the crack follows the line of the maximum principal stresses σ_1 and σ_2 at 45° with respect to the axis.

In brittle materials or in conditions of brittle behavior (three-dimensional states of stress), the micro-defects are formed directly from the discontinuities at the surface or internally without the formation of beach marks and they propagate in the direction perpendicular to the maximum normal stress; the marks are not always visible because the hammering of the surfaces of the crack resulting from the compressive half-cycle, leaves smoothed areas, without marks.

7.12 Influence of the Surface Finishing

The presentation of the stress analysis models for high cycle fatigue phenomenon does not discuss the influence on fatigue of the surface state, since this very important and fundamental aspect is of an empirical nature and is not modeled by mathematical laws.

The surface roughness is a result of micro notches on the surface and the average roughness R_z is the parameter that represents them. The effect of surface condition on the fatigue strength reduction is synthesized in a factor, called surface factor b_s (with value equal to 1 in the case of lapped surfaces). The trend of this factor is shown in Fig. 7.45 as a function of average surface roughness and the static resistance R_m of the material. An extensive presentation of this subject can be found in [22] where a roughness factor and treatment and coating factors are shown and evaluated for different cases.

The experimental analysis guarantees that phenomenological theories of fatigue respect a principle of consistency between slender and massive bodies.

Moreover, phenomenological theories can be established without forcing models against the empirical evidence: e.g., the theory of fatigue at a high number of cycles can not be used to describe the dubious correlations between the number of cycles and stresses of high amplitude. In these cases it is better to abandon this theory.

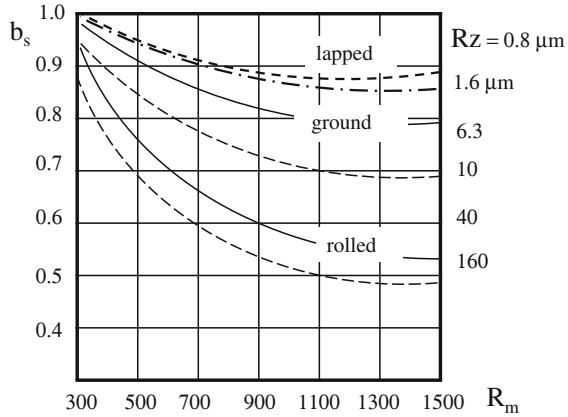
This observation explains the need for a second fatigue theory based on the local deformations that take the place, as control parameter, of the local stress and gradient, concepts defined only in the elastic range.

Another critical aspect of the High Cycle Fatigue theory is the search for a stress limit of the S/N curves for materials that have uncertain horizontal asymptote: for a number of cycles of a higher order of magnitude, this conventional limit may lose significance in establishing a level of safety for the design against fatigue.

7.13 Dixon Method

The determination of the fatigue limit of a component is usually quite a difficult task, as it requires the design and the development of a specific loading device. In particular, it is important that the experimental set-up correctly reproduces the actual load on the component during real service. One consequence is that performing tests on real parts is quite an expensive task, which can be time consuming when a high testing frequency cannot be used or if tests must be continued up to a huge number of cycles at the fatigue limit. From the methodological point of view, some Standards are

Fig. 7.45 Effect of the surface state on fatigue strength in high cycle fatigue



available and can be followed for the design of the experiment and for the statistical processing of the results. More details can be found in Refs. [49–53]. Among these Standards, *ISO 12107 2012* [52] is the most widely used, even if it has the drawback of requiring a minimum of 15 tests to be performed to properly determine a fatigue limit. Therefore, other methodologies may be adopted to achieve the determination of the fatigue limit by a smaller number of tests.

This section deals with this issue: focus is given to the *up & down* method by Dixon [54], with successful applications by Guagliano and Vergani in [55] and in many other References, e.g. [44, 46, 56–59]. The following subsections deal with its application for the conduction of a fatigue campaign and for the processing of the results, along with their statistical assessment.

The *up & down* method can be used for any study concerned with *sensitivity*. In the specific case of fatigue tests, the investigated *sensitivity* must be intended with respect to the load amount, in particular to its range or its amplitude, considering fully reversed or pulsating loads. For loads below the fatigue limit, the tested part has theoretically infinite life. On the other hand, when an incremented load that exceeds the fatigue limit is applied, failure occurs after a finite number of cycles. The *up & down* procedure derives from the experimentations that were conducted on explosives during the second world war.

Let us consider an experimental campaign aimed at the determination of the minimum height, below which a bomb dropped from that height does not explode. The first issue consists in a rough estimation of this height, usually based on previous experimental or theoretical outcomes. Afterwards, a suitable height step, not too high and not too low, must be chosen. The first test is performed dropping the bomb from the approximately estimated minimum height. The height for the second test is then provided by the outcome of the first test: if the bomb explodes, the height must be decreased by one step. Otherwise, if explosion does not take place, a height increase is

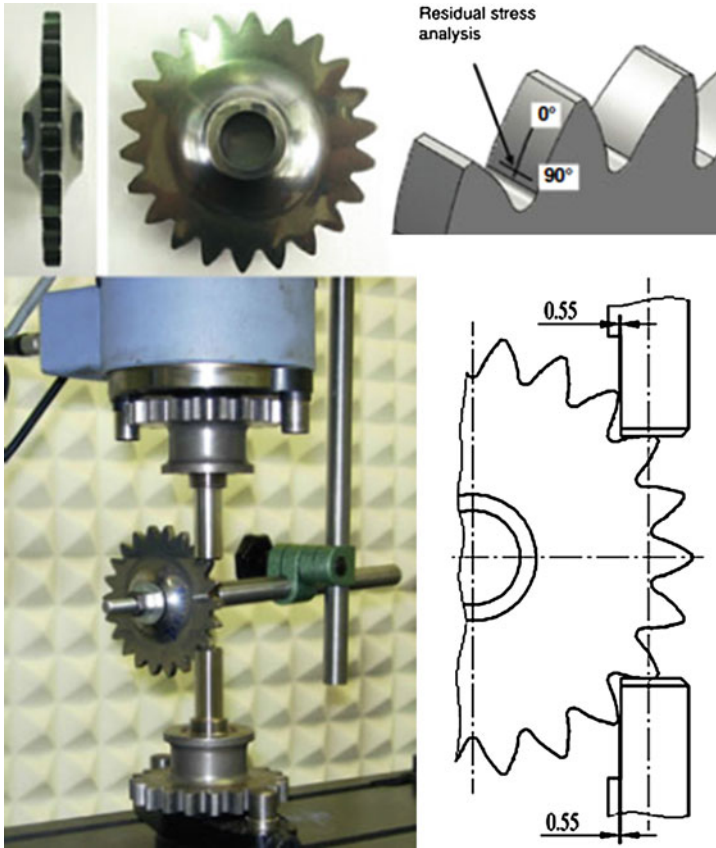


Fig. 7.46 Fatigue Testing on a tooth gear with shot peened steel

suggested. In synthesis, every test must be performed at the upper or at the lower level with respect to the previous one, based on this simple approach. When considering explosive testing, an explosion is traditionally indicated by the symbol *X*, whereas a symbol *O* stands for a misfire.

Back to fatigue testing, the same symbols can be used, where *X* denotes sample failure, whereas *O* stands for not-failure after a suitably selected number of cycles, usually regarded as the *run-out* life. The following example shows how the *UP & DOWN* method can be applied to a fatigue campaign and how the methodology by Dixon [54] can be utilized for the estimation of the fatigue limit and for managing its uncertainty.

7.13.1 Case Study on the Determination of the Local Fatigue Limit of a Mechanical Component

The described task is tackled here with reference to the fatigue limit of gears, considering bending fatigue at the tooth root.

The cyclic load is applied by a pair of twin punches, as shown in Fig. 7.46. The load trend is basically a pulsating one, even if a non-zero minimum load is applied to ensure the correct positioning of the gear with respect to the horizontal surfaces of the punches throughout the test. As a consequence, the load ratio is not exactly zero, but is increased to 0.1.

The tests are performed on a resonant testing machine, operating at a frequency of approximately 110 Hz. The fatigue limit is initially determined in terms of the maximum load (F_{max}) applied by the pair of punches. Afterwards, a simple linear relationship, determined by Finite Element Modeling (FEM), will be introduced to provide the conversion of loads into stresses.

The first step, like in the examples concerned with explosives, consists in the estimation of the most likelihood value for the fatigue limit. In the present case, this value is estimated, based on the material data (high strength steel), on the heating and surface treatments (quenching, carburizing, shot-peening and super-finishing) and on the literature [55, 57, 60–67]. The resulting value, in terms of F_{max} , is 12 kN. Afterwards, an important issue consists in the most proper choice of the load step. This is not an easy choice, as in [54] it is specified that the most suitable value should range between 0.5 and 2σ , where σ stands for the standard deviation affecting the fatigue limit. It is clear that problems arise from the occurrence that the value of the standard deviation can be retrieved only at the end of the campaign. The only available option lies in a reasonable estimation of the standard deviation and therefore of the step value, with a final check of it being within the aforementioned interval. Moreover, it is interesting to observe that the optimal value for the load step is the result of a compromise. A low value would lead to a refined estimation of the fatigue limit; however, it would require many tests, especially considering the many like responses that could be obtained at the beginning of the series.

For instance, if the initial load value is too high, we could have many failure responses before a not-failure occurs.

Conversely, if it is too low, many subsequent tests would lead to not-failure, before having a failure outcome. On the other hand, a higher value is likely to lead to a coarse estimation of the fatigue limit with a fewer number of trials.

In [54] a distinction is made between the overall number of fatigue tests and that of the so-called *nominal tests*. This number, usually regarded as N , can be defined as the total amount of tests (N'), reduced by one less than the number of like responses at the beginning of the series. The response of a fatigue test may be *failure* or *not-failure*, when the test is stopped upon *run-out* (corresponding to the number of cycles at the fatigue limit in the Wöhler curve). When considering specimens in steel, this number usually lies between 10^6 and $5 \cdot 10^6$, to be incremented up to 10^7 for the sake

of safety. For aluminium alloy samples, higher *run-out* values must be set, usually between 10^7 and 10^8 .

The Dixon method [54] makes it possible to roughly estimate the fatigue limit, based on the results of even just two *nominal tests*. However, the recommended amount for the *nominal tests* is at least 6, to obtain a sufficiently reliable result. The uncertainty decreases, as the number of tests increases and becomes generally quite low for 10 *nominal tests* (or more). The proposed formulation for the processing of the experimental data and for fatigue limit computation depends on the number of *nominal tests*. Two cases are possible: N between 2 and 6, or N beyond this amount.

7.13.1.1 Fatigue Limit for N Between 2 and 6

Table 7.14 shows the trend of the responses of the bending fatigue tests. The values in the first column refer to the aforementioned F_{max} . The symbols O and X have the previously explained meaning. Based on previous studies [46, 56, 59], the values of 12 and 0.25 kN were respectively assumed as the initial value of F_{max} and as the load step. According to the *up & down* approach, the failure for the maximum load of 12 kN led to the set-up of a following test, considering a lower F_{max} , 11.75 kN. A further failure led to the further decrease to 11.5 kN of the maximum load. This latest test led to a not-failure response. Consequently, the maximum force was increased by 0.25 kN: the following trial, with the maximum force set at 11.75 kN, led to a not-failure response. Therefore, the load was increased again, and so on

As can be observed, considering the summary of the responses in Table 7.14, the overall number of the trials (N') is 7, whereas the number of like responses at the initial stages is 2. Thus, $N = N' - 2 + 1 = 7 - 2 + 1 = 6$. Considering the two initial trials with the same response, i.e.: failure, only the second one (preceding the trial that led to a not-failure result) must be considered in the computation of the amount of the *nominal tests*. According to [54], the fatigue limit can be estimated as in Eq. 7.45. Let $X_{f_N \leq 6}$ be the fatigue limit at the 50% probability of failure, and X_0 the load level of the last test in the *UP & DOWN* sequence. $X_{f_N \leq 6}$ is determined from that of X_0 , with the addition of an adjustment term, $k \cdot d$. The symbol d stands for the load step, while k is a tabled coefficient that takes the failure/not-failure sequence and the number of initial like responses into account. Its value is yielded by Tables 7.15 and 7.16 [54], where every possible response series is listed.

Table 7.14 Summary of the responses of the fatigue tests ($N \leq 6$)

F_{max} (kN)	Test number						
	1	2	3	4	5	6	7
12.00	X				X		
11.75		X		O		X	
11.50			O				X

Table 7.15 Determination of the value of k , for initial not-failure responses (O) [54]

Second part of series	k for initial like responses ...			
	O	OO	OOO	$OOOO$
X	-0.500	-0.388	-0.378	-0.377
XO	0.842	0.890	0.894	0.894
XX	-0.178	0.000	0.026	0.028
XOO	0.299	0.314	0.315	0.315
XOX	-0.500	-0.439	-0.432	-0.432
XXO	1.000	1.122	1.139	1.140
XXX	0.194	0.449	0.500	0.506
$XOOO$	-0.157	-0.154	-0.154	-0.154
$XOOX$	-0.878	-0.861	-0.860	-0.860
$XOXO$	0.701	0.737	0.741	0.741
$XOXX$	0.084	0.169	0.181	0.182
$XXOO$	0.305	0.372	0.380	0.381
$XXOX$	-0.305	-0.169	-0.144	-0.142
$XXXO$	1.288	1.500	1.544	1.549
$XXXX$	0.555	0.897	0.985	1.000
$XOOOO$	-0.547	-0.547	-0.547	-0.547
$XOOOX$	-1.250	-1.247	-1.246	-1.246
$XOOXO$	0.372	0.380	0.381	0.381
$XOOXX$	-0.169	-0.144	-0.142	-0.142
$XOXOO$	0.022	0.039	0.040	0.040
$XOXOX$	-0.500	-0.458	-0.453	-0.453
$XOXXO$	1.169	1.237	1.247	1.248
$XOXXX$	0.611	0.732	0.756	0.758
$XXOOO$	-0.296	-0.266	-0.263	-0.263
$XXOOX$	-0.831	-0.763	-0.753	-0.752
$XXOXO$	0.831	0.935	0.952	0.954
$XXOXX$	0.296	0.463	0.500	0.504
$XXXOO$	0.500	0.648	0.678	0.681
$XXXOX$	-0.043	0.187	0.244	0.252
$XXXXO$	1.603	1.917	2.000	2.014
$XXXXX$	0.893	1.329	1.465	1.496

$$X_{f_{N \leq 6}} = X_0 + k \cdot d = 11.5 + 0.266 \cdot 0.25 = 11.57 \text{ kN} \tag{7.45}$$

The value of 0.266 was determined for the parameter k , considering the sequence $XX - OOOXX$ and Table 7.16. Finally, the last step consists in the determination of the local fatigue limit at the component reference point. In this case, considering the gear geometry, it was determined at the tooth root. This issue is usually tackled by

Table 7.16 Determination of the value of k , for initial failure responses (X) [54]

Second part of series	- k for initial like responses ...			
	X	XX	XXX	$XXXX$
O	-0.500	-0.388	-0.378	-0.377
OX	0.842	0.890	0.894	0.894
OO	-0.178	0.000	0.026	0.028
OXX	0.299	0.314	0.315	0.315
OXO	-0.500	-0.439	-0.432	-0.432
OOX	1.000	1.122	1.139	1.140
OOO	0.194	0.449	0.500	0.506
$OXXX$	-0.157	-0.154	-0.154	-0.154
$OXXO$	-0.878	-0.861	-0.860	-0.860
$OXOX$	0.701	0.737	0.741	0.741
$OXOO$	0.084	0.169	0.181	0.182
$OOXX$	0.305	0.372	0.380	0.381
$OOXO$	-0.305	-0.169	-0.144	-0.142
$OOOX$	1.288	1.500	1.544	1.549
$OOOO$	0.555	0.897	0.985	1.000
$OXXXX$	-0.547	-0.547	-0.547	-0.547
$OXXXO$	-1.250	-1.247	-1.246	-1.246
$OXXOX$	0.372	0.380	0.381	0.381
$OXXOO$	-0.169	-0.144	-0.142	-0.142
$OXOXX$	0.022	0.039	0.040	0.040
$OXOXO$	-0.500	-0.458	-0.453	-0.453
$OXOOX$	1.169	1.237	1.247	1.248
$OXOOO$	0.611	0.732	0.756	0.758
$OOXXX$	-0.296	-0.266	-0.263	-0.263
$OOXXO$	-0.831	-0.763	-0.753	-0.752
$OOXOX$	0.831	0.935	0.952	0.954
$OOXOO$	0.296	0.463	0.500	0.504
$OOOXX$	0.500	0.648	0.678	0.681
$OOOXO$	-0.043	0.187	0.244	0.252
$OOOOX$	1.603	1.917	2.000	2.014
$OOOOO$	0.893	1.329	1.465	1.496

FEM, in order to find a relationship between the applied load and the corresponding locally generated stress. For instance, in the described application, one-eighth of the gear (a quarter cut along a plane perpendicular to the gear axis) was meshed by solid eight-node linear brick elements and constrained for a free rotation around its axis [46, 56, 59].

The FEM simulation included also one infinitely rigid punch, transmitting a force to the tooth face. Further details are provided in Fig. 7.47. Numerical modelling in the linear-elastic field led to the simple linear relationship in Eq. 7.46, based on the occurrence that a 10 kN load transmitted by the pair of punches generates a stress of 1115 MPa at the tooth root.

The local fatigue limit corresponding to the previously determined (Eq. 7.45) load of 11.57 kN is indicated as $\sigma_{f_N \leq 6}$ (cycle peak in terms of local stress) in Eq. 7.46

$$\sigma_{f_N \leq 6} = \frac{1115}{10} \cdot 11.57 = 1290 \text{ MPa} \tag{7.46}$$

7.13.1.2 Fatigue Limit for N Exceeding 6

The experimental campaign described in Sect. 7.13.1.1 was continued, proceeding with the *up & down* series. Following the failure response for the maximum force of 11.5 kN, the related value was decreased to 11.25 kN. A summary of all the responses is contained in Table 7.17.

Therefore, considering this augmented campaign, the overall number of trials is 10, with 2 like responses at the beginning. Thus, the number of *nominal tests* is: $N = N' - 2 + 1 = 10 - 2 + 1 = 9$, for the full sequence $XX - O O X X X O O X$.

According to [54], the fatigue limit $X_{f_N > 6}$ at the 50% probability of failure for N exceeding 6 must be computed as in Eq. 7.47.

$$\begin{aligned} X_{f_N > 6} &= \frac{\sum_{i=1}^N X_i}{N} + \frac{d \cdot (A+C)}{N} \\ &= \frac{12+11.75 \cdot 4+11.5 \cdot 3+11.25}{9} + \frac{0.25 \cdot (-1.53-0.16)}{9} = 11.59 \text{ kN} \end{aligned} \tag{7.47}$$

In Eq. 7.47 the symbol X_i stands for the load levels of the nominal tests, whereas A and C are tabled coefficients, related to the failure/not-failure sequence and to the number of like responses at the beginning of the series.

They can be determined, based on the data in Tables 7.17, 7.18 and 7.19 [54], where n_X and n_O stand respectively for the number of *nominal tests* with failure and with not-failure responses.

In the present case, we have $n_X - n_O = 5 - 4 = 1$ and, considering the initial failure events, the data in Table 7.19 must be used. It can be observed that the estimated value for the fatigue limit, $X_{f_N > 6} = 11.57$ kN, is very close to that determined in the previous Paragraph ($X_{f_N \leq 6} = 11.59$ kN).

The local fatigue limit is finally computed by the same linear relationship described above. Its value, corresponding to the previously determined maximum load of 11.59 kN (Eq. 7.47) is determined in Eq. 7.48 and indicated by the symbol $\sigma_{f_N > 6}$ (cycle peak in terms of local stress).

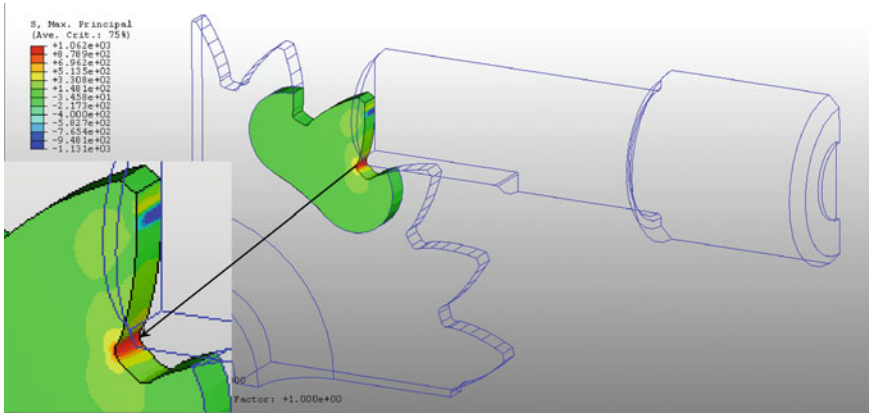


Fig. 7.47 Scheme of numerical stress analysis of the tooth gear

Table 7.17 Resume of the responses of the fatigue tests ($N > 6$)

F_{max} [kN]	Test number										n_O	n_X	
	1	2	3	4	5	6	7	8	9	10			
12.00	X				X							0	1
11.75		X		O		X				X		1	3
11.50			O				X		O			2	1
11.25								O				1	0

Table 7.18 Determination of the parameters A and C , for initial not-failure responses (O) [54]

$n_O - n_X$	A	C for initial like responses ...				
		O	OO	OOO	$OOOO$	$OOOOO$
5	10.8	0	0	0	0	0
4	7.72	0	0	0	0	0
3	5.22	0	0.03	0.03	0.03	0.03
2	3.20	0	0.10	0.10	0.10	0.10
1	1.53	0	0.16	0.17	0.17	0.17
0	0	0	0.44	0.48	0.48	0.48
-1	-1.55	0	0.55	0.65	0.65	0.65
-2	-3.30	0	1.14	1.36	1.38	1.38
-3	-5.22	0	1.77	2.16	2.22	2.22
-4	-7.55	0	2.48	3.36	3.52	3.56
-5	-10.3	0	3.5	4.8	5.2	5.3

Table 7.19 Determination of the parameters A and C , for initial failure responses (X) [54]

$n_X - n_O$	$-A$	$-C$ for initial like responses ...				
		X	XX	XXX	$XXXX$	$XXXXX$
5	10.8	0	0	0	0	0
4	7.72	0	0	0	0	0
3	5.22	0	0.03	0.03	0.03	0.03
2	3.20	0	0.10	0.10	0.10	0.10
1	1.53	0	0.16	0.17	0.17	0.17
0	0	0	0.44	0.48	0.48	0.48
-1	-1.55	0	0.55	0.65	0.65	0.65
-2	-3.30	0	1.14	1.36	1.38	1.38
-3	-5.22	0	1.77	2.16	2.22	2.22
-4	-7.55	0	2.48	3.36	3.52	3.56
-5	-10.3	0	3.5	4.8	5.2	5.3

$$\sigma_{f_{N>6}} = \frac{1115}{10} \cdot 11.59 = 1292 \text{ MPa} \tag{7.48}$$

7.13.2 Statistical Analysis of the Results

The previous Sections have dealt with the estimation of the fatigue limits, based on the experimental results. However, a statistical analysis is required to determine the reliability of the retrieved values. Moreover, an important issue consists in the estimation of the confidence interval to be applied to the retrieved limit, depending on the scattering of the results that led to its computation.

The recommended procedure for the statistical analysis is explained here with reference to the case in Sect. 7.13.1.2 (with 9 nominal tests) due to the greater statistical evidence of the results. However, the data can be processed in the same way for whatever number of *nominal tests*, below or above 6.

The first step consists in the assumption that the critical levels of stress or load, considering fatigue failure, are normally distributed. Let μ and σ^2 be the mean value and the variance of the corresponding stochastic distribution. Their *real* values could be determined only if a very largely sized (theoretically infinite) data population were available. For a limited size database, their maximum likelihood estimates can be processed, based on the mean value and on the variance of the available test results. Indicating the stress or load levels by the letter y , the symbol \bar{y} is used here to denote their mean values and s_y^2 to indicate their variance.

One important issue consists in the remark that only failure or only not-failure data must be used in the following computations, depending on which has the smaller total. In other words, first of all, it is necessary to establish if there were fewer

failure (symbol X) or fewer not-failure (symbol O) responses. In the experiment described in Sect. 7.13.1.2 there were fewer not-failures (4) rather than failures (5) (see Table 7.17), therefore, only the not-failure data must be used. Let n stand for the smaller total of events (in this case: $n = n_O =$ number of not-failure events), and let $n_0, n_1, n_2, \dots, n_i, \dots, n_k$ denote the frequencies of the less frequent event, where n_0 is related to the lowest level and n_k corresponds to the highest. It must be: $\sum_{i=0}^k n_i = n$. The following equations provide details on the computation of the mean \bar{y} , of the mean estimate μ , and of the standard deviation σ (Table 7.18).

Considering the results in Sect. 7.13.1.2, the not-failure events O occurred at 11.25, 11.5 and 11.75 kN, which are respectively levels 0, 1, and 2. Therefore, the corresponding frequencies are: $n_0 = 1, n_1 = 2$ and $n_2 = 1$, with $n = 4$. The mean \bar{y} is yielded by Eq. 7.49.

$$\bar{y} = \frac{\sum_{i=0}^k n_i y_i}{\sum_{i=0}^k n_i} = \frac{n_0 y_0 + n_1 y_1 + n_2 y_2}{n_0 + n_1 + n_2} = \frac{11.25 + 2 \cdot 11.5 + 11.75}{1 + 2 + 1} = 11.5 \text{ kN} \quad (7.49)$$

The estimate for μ is given by the previously computed fatigue limit $X_{f_{N>6}}$.

The estimation of the standard deviation σ is achieved by Eq. 7.50. This is an empirical relationship, having the particular feature of providing the estimate of σ as a linear function of the variance s_y^2 , whereas it could be expected that $\sigma^2 \propto s_y^2$. As a first step, the variance term must be computed, as in Eq. 7.50.

$$s_y^2 = \frac{\sum_{i=0}^k [n_i \cdot (y_i - \bar{y})^2]}{\sum_{i=0}^k (n_i) - 1} = 0.042 \text{ kN}^2 \quad (7.50)$$

$$\sigma \approx 1.62 \cdot d \cdot \left(\frac{s_y^2}{d^2} + 0.029 \right) = 1.62 \cdot 0.25 \cdot \left(\frac{0.042}{0.25^2} + 0.029 \right) = 0.28 \text{ kN} \quad (7.51)$$

The expression in Eq. 7.51 may be regarded as an acceptable estimation of the standard deviation σ , when the ratio s_y^2/d^2 is higher than 0.3. Otherwise, it becomes quite rough in the determination of σ on the basis of s_y^2 , when the aforementioned ratio is much lower than 0.3.

Considering the linear relationship between the applied load and the generated local stress determined in Sect. 7.13.1.1, a standard deviation of 0.28 kN corresponds to 31 MPa.

An important issue at this stage consists in the computation of the ratio between the step entity and the standard deviation, i.e.: d/σ . In this case we have: $d/\sigma = 0.89$:

the retrieved value is well contained in the interval between 0.5 and 2, which confirms the good choice of the load step and makes sure of the reliability of the determined fatigue limit.

As a final step, the standard error $\bar{\sigma}$ must be computed. According to [54], considering a normally distributed data population, $\bar{\sigma}$ can be approximated as $\sigma \cdot \sqrt{2/N}$. N indicates again the overall number of the nominal tests and the ratio $\frac{N}{2}$ is very close to n . In the present case, the standard error, which accounts for the scattering of the results in Sect. 7.13.1.2, Table 7.17, is 0.13 kN. This outcome makes it possible to finally compute a confidence interval to be applied to the fatigue limit $X_{f_{-N>6}}$ computed in Eq. 7.47. The procedure for the computation of the 95 % confidence interval is shown in Eq. 7.52.

$$X_{f_{-N>6}} \pm 1.96 \cdot \bar{\sigma} = 11.59 \pm 1.96 \cdot 0.13 = [11.33; 11.85] \quad (7.52)$$

Considering again the linear relationship determined in Sect. 7.13.1.1, the standard error in terms of stress is 14.5 MPa. As a consequence, the 95 % confidence interval for local stresses is between 1263 and 1321 MPa. The *real* value of the local fatigue limit is likely to be contained within this interval.

References

1. ASTM E1823-13 (2013) Standard terminology relating to fatigue and fracture testing. ASTM International, West Conshohocken, PA, USA. <http://www.astm.org/Standards/E1823.htm>
2. Budynas R, Nisbett K (2014) Shigley's mechanical engineering design, vol 1, 10th edn. McGraw-Hill Series in Mechanical Engineering. McGraw-Hill, Milano. ISBN-13: 978-0073398204
3. Haibach E (2002) Betriebsfestigkeit: Verfahren und Daten zur Bauteilberechnung. Springer, Berlin
4. Niemann G, Winter H, Höhn BR (2001) Maschinenelementen, vol 1. Springer, Berlin
5. Stephens RI, Fatemi A, Stephens RR, Fuchs HO (2002) Metal fatigue in engineering, vol 1, 2002nd edn. Wiley, New York
6. Smith R (2004) Railway technology—the last 50 years and future prospects. JRTR: Jpn Railw Trans Rev 27:16–24
7. Freddi A (2012) Contributo alla storia della fatica dei materiali dalle origini agli albori del XX secolo. In: E Mesini DM (ed) Scienza e Tecnica nel Settecento e nell'Ottocento, vol 1, Clueb, Via Marsala 31 40126, Bologna, pp 255–284
8. Wöhler A (1858) Bericht über die Versuche, welche auf der K. Niederschlesisch. Eisenbahn mit Apparaten zum Messen der Biegung und Verdrehung von Eisen-Achsen während des Fahrt, angestellt wurden. Z Bauw 8:642–651
9. Gradassi AR (2001) Progetto e costruzione di una cella di carico per la misurazione di forze in attrezzature sportive (scarpone da sci). MS thesis, University of Bologna
10. Jung P (1975) Zur messtechnik an skiern. Messtechnische Briefe 11(2):34–37
11. Jung P (1975) Zur messtechnik an skiern. Messtechnische Briefe 11(3):59–62
12. Wirsching P, Ortiz K (1992) Reliability methods in mechanical and structural design. In: 13th annual seminar and workshop, Arizona 85721, The University of Arizona, College of Engineering, Aerospace, Mechanical Department, Tucson
13. Olmi G (2014) An experimental investigation on a crack propagating from a geartrain housing in an asphalt milling machine. Eng Fail Anal 38:38–48

14. Olmi G (In press) Failure of the chassis of roller skates for agonistic figure skating. *Case Stud Eng Fail Anal*. doi:[10.1016/j.csefa.2014.08.002](https://doi.org/10.1016/j.csefa.2014.08.002)
15. Olmi G, Freddi A, Croccolo D (2008) In-field measurement of forces and deformations at the rear end of a motorcycle and structural optimisation: experimental-numerical approach aimed at structural optimization. *Strain* 44(6):453–461
16. Pellicoli G, Petrone N (2010) Experimental strain analysis of the mast of a 420 sailboat during sailing. *Strain* 46:482–492
17. Petrone N (2011) Field acquisition of ski boots flexural behaviour for the correct definition of standard laboratory tests. In: *Proceedings 19th international conference on Ski Trauma and safety*, Keystone, Colorado, USA
18. Scott NW, Yoneyama T, Kagawa H, Osada K (2007) Measurement of ski snow-pressure profiles. *Sports Eng* 10:145–156
19. Xu YL, Chen ZW, Xia Y (2012) Fatigue assessment of multi-loading suspension bridges using continuum damage model. *Int J Fatigue* 40(0):27–35. doi:[10.1016/j.ijfatigue.2012.01.015](https://doi.org/10.1016/j.ijfatigue.2012.01.015)
20. Ye XW, Ni YQ, Wong KY, Ko JM (2012) Statistical analysis of stress spectra for fatigue life assessment of steel bridges with structural health monitoring data. *Eng Struct* 45(0):166–176, doi:[10.1016/j.engstruct.2012.06.016](https://doi.org/10.1016/j.engstruct.2012.06.016), <http://www.sciencedirect.com/science/article/pii/S0141029612003197>
21. Beretta S, Cheli F et al. (1999) Dynamic simulation and estimation of stress spectra for railways components. In: *Proceedings 4th ADAMS/RAIL users, 4th ADAMS/RAIL users conference*, Utrecht 1999, vol 1. <http://web.mscsoftware.com/support/library/conf/adams/rail/pdf/milano.pdf>
22. Haibach E (2003) FKM-guideline: analytical strength assessment of components in mechanical engineering, 5th edn. VDMA, Frankfurt
23. Albert WAJ (1838) Über Treibseile am Harz. *Archive für Mineralogie Geognosie Bergbau und Hüttenkunde* 10:215–234
24. Timoshenko S (1953) *History of the development of strength of materials in Russia*, vol 1, 1st edn. Accademia Nazionale dei Lincei, Roma
25. Rumul resonance machine (2010). http://www.rumul.ch/400_service.php?lang_choose=1
26. Zwick machine (2010). <http://www.zwickroell.it/it.html>
27. Eichlseder W (2002) Fatigue analysis by local stress concept based on finite element results. *Comput Struct* 80:2109–2113
28. Hück M, Thrainer L, Schütz W (1981) Berechnung von Wöhlerlinien für Bauteile aus Stahl, Stahlguss und Grauguss-Synthetische Wöhlerlinien. VDEh, Düsseldorf, Germany, bericht Nr. ABF 11 (2. überarbeitete Fassung)
29. Dowling NE (2012) *Mechanical behavior of materials*, vol 1, 4th edn. Prentice Hall, Upper Saddle River. ISBN-13: 978-0131395060, ISBN-10: 0131395068
30. Taylor D (2007) *The theory of critical distances—a new perspective in fracture mechanics*. Elsevier Science B.V., Amsterdam
31. Lazzarin P, Tovo R, Meneghetti G (1997) Fatigue crack initiation and propagation phases near notches in metals with low notch sensitivity. *Int J Fatigue* 19:647–657
32. Liu CD, Bassim MN, Lawrence SS (1995) Dependence of the fatigue limit of rail steels on stress intensity factor near inclusions. *Eng Fract Mech* 50(2):301–307, doi:[10.1016/0013-7944\(94\)00197-P](https://doi.org/10.1016/0013-7944(94)00197-P), <http://www.sciencedirect.com/science/article/pii/001379449400197P>
33. Kuske A, Robertson G (1974) *Photoelastic stress analysis*, vol 1, 1st edn. Wiley, London
34. Siebel E, Stieler M (1955) Ungleichformige spannungsverteilung bei schwingender beanspruchung. *VDI-Z* 97:121–126
35. Stieler M (1954) *Untersuchungen Über die Dauerschwingfestigkeit metallischer Bauteile bei Raumtemperatur*. Dissertation, TU Stuttgart
36. Siebel E, Bussmann KH (1948) Das kerbproblem bei schwingender beanspruchung. *Technik* 3:249–252
37. Eichlseder W (2003) Lebensdauervorhersage auf Basis von Finite Elemente Ergebnissen (Fatigue life prediction based on finite element results). *Werkstofftech* 34:843–849
38. Eichlseder W (2005) *Betriebfestigkeit*, vol. 1. Montan Universitaet Leoben

39. Eichlseder W, Leitner H (2002) Influence of stress gradient on s/n-curve. In: Proceedings of the new trends in fracture and fatigue congress, Metz, France, 8–9 April 2002
40. Gänser HP, Gódor I, Leitner H, Eichlseder W (2006) Enhanced fatigue life by mechanical surface treatments—experiment and simulation. In: Proceedings 16th European conference of fracture, Alexandroupolis, Greece
41. Leitner H, Eichlseder W, Gódor I, Waggermayer M, Hinteregger C (2005) Increased fatigue limit of gear wheels by a combination of case hardening and shot peening. In: Proceedings RELMAS 2005 (6th international conference), St Petersburg, Russia
42. Leitner H, Gänser HP, Eichlseder W (2007) Oberflächennachbehandlung durch kugelstrahlen und festwalzen-mechanismen, modellierung, methoden. *Materialprüfung* 49:408–413
43. Reggiani B (2007) Simulation models in biomechanics and experimental mechanics. PhD thesis, University of Bologna, available as PDF file at <http://www.ingegneriaindustriale.unibo.it/it/attivita-didattica/download-tesi-di-dottorato>
44. Olmi G, Freddi A (2013) A new method for modelling the support effect under rotating bending fatigue: application to Ti-6Al-4V alloy, with and without shot peening. *Fatigue Fract Eng Mater Struct* 36(10):981–993
45. Pilkey WD, Pilkey DF (2008) Peterson's stress concentration factors, 3rd edn. Wiley, Hoboken
46. Comandini M (2008) Progettazione con metodologie avanzate di organi di macchine sollecitati a fatica, corpi non saintvenantiani. PhD thesis, University of Bologna, available as PDF file at <http://www.ingegneriaindustriale.unibo.it/it/attivita-didattica/download-tesi-di-dottorato>
47. Gambini L (2007) Verifica a fatica di un braccio di escavatore con le teorie della tensione locale, della deformazione locale e relativo confronto con le principali normative. Master thesis, University of Bologna. <https://www.leonardo.gambini@libero.it>
48. Rankine WJM (1842) On the causes of the unexpected breakage of the journals of railway axles, and on the means of preventing such accidents by observing the law of continuity in their. In: Institution of Civil Engineers, minutes of proceedings, vol 1, pp 105–108
49. ASTM E468-11 (2011) Standard practice for presentation of constant amplitude fatigue test results for metallic materials. ASTM International, West Conshohocken, PA, USA
50. ASTM E739-10 (2010) Standards practice for statistical analysis of linear or linearized stress-life (S-N) and strain-life (ϵ -N) fatigue data. ASTM International, West Conshohocken, PA, USA
51. Barsom JM, Rolfe ST (1999) Fracture and fatigue control in structures: applications of fracture mechanics. ASTM manual series, vol 41. ASTM International, West Conshohocken, PA, USA. ISBN: 9780803120822
52. ISO 12107:2012 (2012) Metallic materials—fatigue testing—statistical planning and analysis of data. International Organization for Standardization (ISO), Geneva, Switzerland
53. Stephens RI, Fatemi A, Stephens RR, Fuchs HO (2000) *Met Fatigue Eng*, 2nd edn. Wiley, Hoboken
54. Dixon WJ, Massey FJ Jr (1983) *Introduction to statistical analysis*, vol 1. McGraw-Hill, New York
55. Guagliano M, Vergani L (2004) An approach for prediction of fatigue strength of shot peened components. *Eng Fract Mech* 71:501–512
56. Comandini M, Olmi G, Freddi A (2007) Fatigue performance of shot-peened gears investigated by experimental and numerical methods. *Trans Famena* 31–2:1–10
57. Crococolo D, Cristofolini L, Freddi A, Bandini M (2002) Fatigue strength of shot peened nitrided steel: optimization of process parameters by means of design of experiment. *Fatigue Fract Eng Mater Struct* 25:695–707
58. Olmi G, Freddi A (2006) Shot peening technique for fatigue improvement of high strength steel gears. In: Proceedings 5th international conference of the Croatian society of mechanics (ICCSM), Trgir/Split, Croatia, paper number: 153
59. Olmi G, Comandini M, Freddi A (2010) Fatigue on shot-peened gears: experimentation, simulation and sensitivity analyses. *Strain* 46(4):382–395
60. Benedetti M, Fontanari V, Höhn BR, Oster P, Tobie T (2002) Influence of shot peening on bending tooth fatigue limit of case hardened gears. *Int J Fatigue* 24:1127–1136

61. Freddi A, Veschi D, Bandini M, Giovani G (1997) Design of experiments to investigate residual stresses and fatigue life improvement by a surface treatment. *Fatigue Fract Eng Mater Struct* 20(8):1147–1157
62. Guagliano M, Vergani L, Bandini M, Gili F (1999) An approach to relate the shot-peening parameters to the induced residual stresses. In: *Proceedings ICSP7, Warsaw, Poland*
63. Guagliano M, Riva E, Guidetti M (2002) Contact fatigue failure analysis of shot-peened gears. *Eng Fail Anal* 9:147–158
64. Herzog R, Zinn W, Sholtes B, Wohlfahrt H (1996) The significance of Almen intensity for the generation of shot peening residual stresses. In: *Proceedings of the VI international conference on shot-peening*, pp 270–281
65. Mann JY (1967) *Fatigue of materials—an introductory text*. Melbourne University Press, Melbourne
66. Pariente IF, Guagliano M (2008) About the role of residual stresses and surface work hardening on fatigue δk_{th} of a nitrided and shot peened low-alloy steel. *Surf Coat Technol* 22:3072–3080
67. Shaw BA, Aylott C, O'Hara P, Brimble K (2003) The role of residual stress on the fatigue strength of high performance gearing. *Int J Fatigue* 25:1279–1283

Chapter 8

Local Strain Models for Variable Loads

Abstract In many practical applications of fatigue loadings with low amplitudes a number of higher load cycles are applied that overcome the threshold of the material yield strength, even if only locally. This circumstance is not taken into account by the classic theory of fatigue at a high number of cycles that assumes elastic states of stress at any point of a loaded body and a limited micro-plasticity at critical points. In the field of high load cycles, damage due to local plasticity is instead the governing phenomenon. Simulation models are developed based on a *strain-controlled* approach, especially in the cases of complex (i.e. not constant but time-varying) loads amplitudes. This local nonlinear cyclic behavior leads to a damage that occurs for a limited number of cycles $< 5 \times 10^4$. This numerical limit is conventional, based on experience and therefore subjected to different estimations by different authors. A similar convention emerges in establishing the test conclusion when the first visible crack appears (an event subjected to interpretation).

The fatigue process in real structures or in samples during laboratory tests, shows a damage evolution at different characteristic stages, see also Fig. 7.37 in Chap. 7:

1. A initial stage characterized by a number of load cycles (hereinafter also referred to a portion of component life), up to the nucleation of initial barely visible defects of the order of magnitude of $0.5 \div 1$ mm that are limited to the first material grains.
2. A stage of propagation up to a critical crack size. When this value is reached, a complete static rupture of the specimen occurs.

While the first stage is not practically discernible in High Cycle Fatigue due to the rapid cycles accumulation, so that only the final rupture is considered a reliable parameter for a fatigue limit evaluation, in the case of *Low Cycle Fatigue*, the first limit must be identified by the record of the cycles corresponding to a crack nucleation, even if with a certain approximation and conventionality. The number of load cycles until the formation of the first visible defects may constitute a

significant part of the total component life: in some design methodologies only the life up to the formation of the first defect is considered useful.

The second phase relative to propagation of macroscopic defects, is the classic fatigue propagation phase. The propagation of a crack until it reaches a critical dimension may, in some cases, constitute a significant portion or even the only acceptable period of life.

This stage is interpreted by the *Fracture Mechanics* models, see Chap. 9.

8.1 Experimental Analysis of Local Deformation

A Low Cycle Fatigue Theory must take into account all the following limitations:

- The developed models must restrict the analysis to a *local* non-linear behavior of small areas at high stress and strain concentrations, since the phenomenon is localized, while the remaining part of the body still exhibits elastic behavior [1].
- The uncertainty on the resolution of the defect visibility at the test end (even if validated by *NDT* techniques and by measurements on the compliance of the body), must overcome the limit of a personal judgment, such as to allow a comparison between results obtained in different laboratories on identical materials.
- The theory must be able to handle the possible variation of the load amplitudes over time. This is a reasonable problem when stress amplitudes remain in elastic field and can be collected in stress spectra, Chap. 7 and [2], while when cycle amplitudes exceed the elastic limit, the local damage depends not only on the parameters of the cycle but also on the entire previous load history. In this case, the actual sequence of the load cycles must be taken into account, since material (over the elastic limit) maintains a *memory* of the previous loads and they can not be collected in spectra, insensitive to the sequence.
- As it was seen in Chap. 7, under variable amplitude loading the definition of load *cycles* can become ambiguous and in the new Low Cycle fatigue theory it is preferable to define load *reversals*. At constant amplitude one cycle consists of two reversals.
- Finally, it is necessary to establish a theory that takes into account high load cycles, for which material locally exceeds the elastic limit [3, 4] but it is also able to consider loads with low amplitudes with stresses that remain in the elastic field [5, 6].

8.2 Laboratory Testing

Figure 8.1 shows two commercial servo-hydraulic universal testing machines used for low but also for high cycle fatigue tests; the presentation in this chapter is justified by their specific capability of *controlled deformation* by a sensor mounted on the



Fig. 8.1 Example of universal testing machines of 250 and 500 kN capacity (INSTRON Ltd.)

specimen itself. This is a special device on the specimen or a simple strain gage placed at a reference point. The machines also employ high stiffness load cells of the kind shown in Chap. 2.

They work under ideal conditions for frequencies in the interval $0.01 \div 5$ Hz, while they are not the most suitable for high-frequencies due to the strain controlling difficulty. Large deformations, typical of low cycle fatigue, would also be incompatible with high frequencies due to the overheating of the specimen and other serious drawbacks.¹

In Fig. 8.2, in order to perform tests in special environments, e.g. at low temperatures in cryogenic conditions, a frame was designed with a hydraulic cylinder mounted on the upper cross-head and a tank for liquid nitrogen to the lower cross member. This avoids a possible damage by a leakage of nitrogen on the cylinder of the classic testing machines.

On the basis of these premises of a physical character, it is necessary to develop several models for the simulation of each of the following steps [7–9]:

- Simulation of cyclic behavior in the elastic-plastic field.
- Simulation of geometrical discontinuities beyond the elastic limit.
- Simulation of the strain-controlled fatigue life.
- Simulation of loads sequences of variable amplitudes.
- Simulation of the damage for each load cycle and of the accumulated damage for the whole load history.

¹ Details in A.S.T.M. standardized procedure to determine the fatigue curve in deformation control for lives at low number of cycles, in accordance with *ASTM—Standard E606–12*.

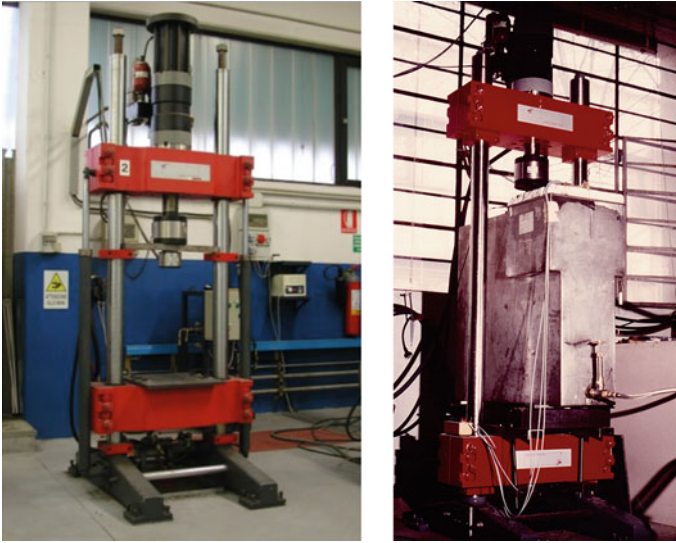


Fig. 8.2 Special self-manufactured testing apparatus with the hydraulic cylinder on the upper cross-head for 250 kN tests, without and with a tank for liquid nitrogen for low temperature and low cycles fatigue tests

8.3 Strain-Controlled Versus Stress-Controlled Tests

Let us consider a sample of homogeneous and isotropic material and analyze the relationship between stress and strain at a point. The material shows an asymmetry of behavior if, after loading in tension beyond the yield point, the specimen is loaded in compression Fig. 8.3a, since the yield strength in compression is affected by the previous tensile load, *Bauschinger effect*.

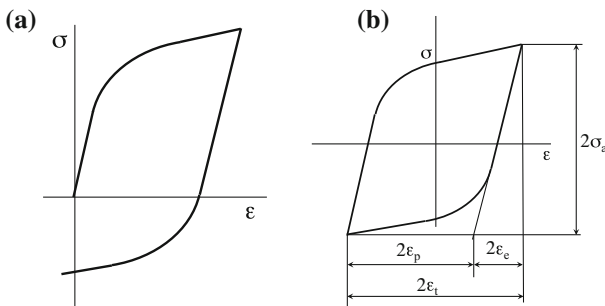


Fig. 8.3 **a** Stress/strain diagram for a compression after tension (Bauschinger effect) and **b** a hysteresis cycle. Symbols σ_a = stress amplitude, ε_t = total strain amplitude, ε_e = elastic strain amplitude, ε_p = plastic strain amplitude

If, however, the specimen is reloaded several times, maintaining the interval of the applied strain constant, hysteresis cycles are generated that tend to stabilize on a final shape after a certain number of load cycles, Fig. 8.3b.

8.3.1 Repeated Loading in Stress-Controlled Tests

Figure 8.4a, b shows the material response for stress-controlled axial load. Differences of behavior of the stabilized loop depend on the state of the material before the test: a soft materials show cyclic hardening and cold-worked materials show cyclic softening.

8.3.2 Repeated Loading in Strain-Controlled Tests

If the specimen is loaded with a constant amplitude of the applied strain, the behavior of the material is characterized by hysteresis cycles that tend to stabilize after a certain number of load cycles, reducing or increasing the value of the maximum cycle stress; a soft material shows cyclic hardening and vice-versa a cold-worked material shows cyclic softening Fig. 8.5. When the specimen is loaded and unloaded several times while keeping the strain amplitude at constant value, a new stress/strain curve can be defined that synthetically represents the cyclic behavior of the material. It is obtained with the following procedure:

- A series of cycles of constant strain amplitude are applied to the sample; the hysteresis cycle tends to stabilize on a final configuration.

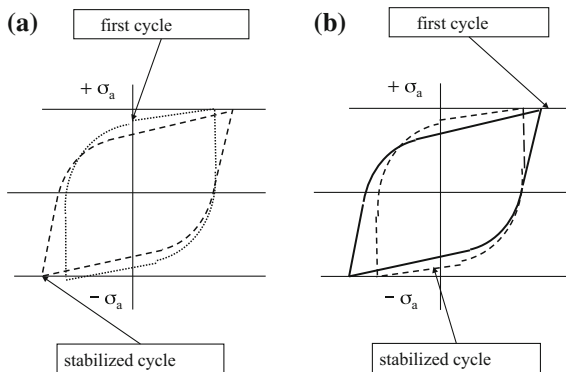


Fig. 8.4 Stress-controlled axial load: **a** cyclic hardening of a material initially annealed and **b** cyclic softening of a material initially cold-worked

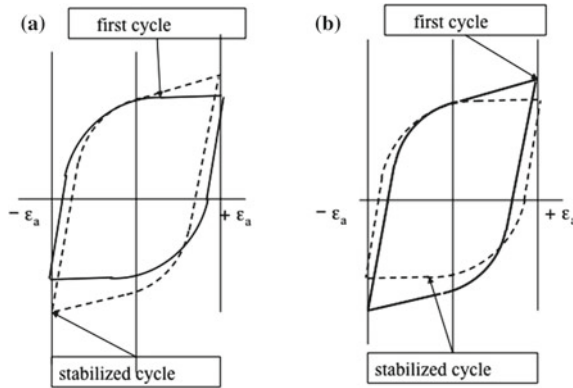


Fig. 8.5 Stress-strain behavior for strain-controlled axial load: **a** cyclic hardening of a material initially annealed and **b** cyclic softening of a material initially cold-worked

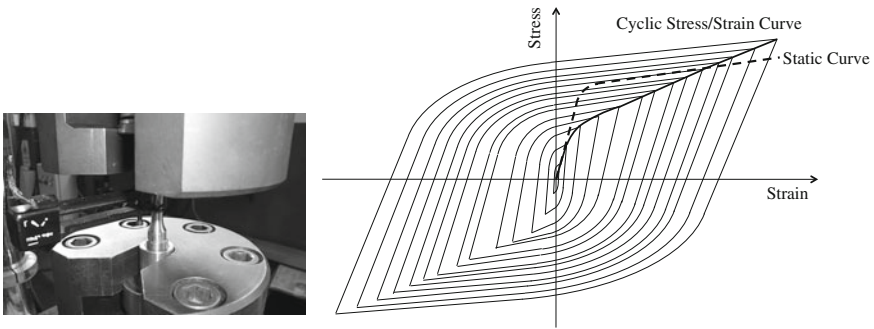


Fig. 8.6 Device for strain control (Instron Ltd.), mounted on the specimen and envelope of the cusps of the stabilized hysteresis loops for building the cyclic curve

- Strain amplitudes are increased by a discrete quantity and a second hysteresis loop is obtained and stabilized.²
- The procedure is repeated for a number of increased strain amplitudes.

The cusps envelope of the cycles of stabilized hysteresis cycles is named the *cyclic stress/strain curve* of a material.

Figure 8.6 on the left shows a device named *extensometer*, mounted on a cylindrical specimen. The *axial extensometer* controls the deformation through the displacement measurement between two points on the surface of the specimen, at a certain distance one to another. Its simple scheme is presented in Chap. 9, Sect. 9.6.1.2. Possible alternatives to the use of an *axial extensometer* like that in Fig. 8.6 consist in the application of a *transverse extensometer* (also known as *diametral clip gage*)

² Using the same specimen could produce results invalidated by the material training by the previous cycles. The best practice is the use of a new specimen for each test.

[10–12] or in the adoption of an optical extensometer [13–17]. However, these devices are rarely utilized for closed loop control of the testing machine due to insufficient measurement accuracy and reliability [18]. A further alternative, which is suitable for closed-loop strain-controlled testing of both uniform gage and hourglass specimens, is described in [18].

The same Figure on the right shows the hysteresis loops of several constant strain amplitudes and the cyclic stress/strain curve obtained by the envelope of the cusps of hysteresis cycles (continuous line). It is compared with the static stress/strain curve (dotted line).³

8.4 Model for Material Behavior

The behavior of the material in the elasto-plastic field (the constitutive law between stress and strain) is considered independent of the presence or absence of a notch and can therefore be determined with the reference to smooth specimens. The simplest test is on a cylindrical specimen loaded in tension/compression, in which the constitutive law concerns a single stress and a single strain parameter.

In the case of multi-axial states of stress varying with the time, the problem becomes much more difficult and a general method to deal with it is not yet known. However, a considerable effort was made for classify and regularize the cases that can be approached and it is necessary to refer to the literature to find possible solutions. Multi-axial stresses varying with time have to be classified as follows [2]:

- proportional stresses,
- synchronous stresses,
- non-proportional stresses.

Here, the presentation will be limited to the most simple assumption of the first case of proportional stresses when all stresses at a reference point of an element vary proportionally to each other in amplitudes and also in mean value, or at least of synchronous stresses in which proportionality is limited to the stress amplitudes. This hypothesis can be accepted, due to the prominent importance of the amplitudes compared to the mean values, with regard to the influence on the fatigue strength of a component.

In these cases it is possible to approximate the behavior of multi-axial stress states through equivalent stress and strain.⁴

³ A difficulty could occur in the identification of the stabilized hysteresis loop. Often, depending on the material and strain level, dynamic stability is not reached, i.e. stress amplitude corresponding to an imposed strain amplitude tends to vary continuously during the test and, in some cases, until the rupture of the specimen. In this case a conventional definition of stabilized cycle is given, as the hysteresis cycle that corresponds to the mid-life of the specimen, i.e. to the half number of cycles to failure.

⁴ In the case of components with states of non-proportional biaxial or triaxial stress (with loads agents not in phase with each other), in order to establish some reasonable estimation of the component strength, the reference specimen for analyzing the cyclic properties of the material must be arranged in such a way as to reproduce the same phase relationships that occur in reality [2, 19].

For the analytic representation of the cyclic curve, the same formula seen for the static case is utilized, but with an update of the parameters that describe the material (i.e. K' and n'):

$$\varepsilon_t = \varepsilon_e + \varepsilon_p = \frac{\sigma}{E} + \left(\frac{\sigma}{K'}\right)^{1/n'} \quad (8.1)$$

In Eq. 8.1 the symbols have the following meaning, referring to Fig. 8.3:

- σ_a = true stress amplitude
- ε_a = total true strain amplitude
- ε_e = true elastic strain amplitude
- ε_p = true plastic strain amplitude
- K' = cyclic strength coefficient
- n' = cyclic strain hardening exponent.

The cyclic curve permits the analytical description of the ascending and descending branches of the hysteresis cycles that have generated it. For the way in which it was obtained, if the lower cusps are overlapped and the rectilinear ascending parts of all the cycles are superimposed, the higher cusps lie on a curve that is homothetic with respect to the cyclic curve in a ratio 2 : 1, i.e. the branches can be described by the same Eq. 8.1 as long as the variables σ and ε are replaced with the new variables $\sigma/2$ and $\varepsilon/2$, Fig. 8.7.

The ascending and descending curves of the hysteresis loops can then be described by the following equations, using the amplitudes of stress and strain for referring

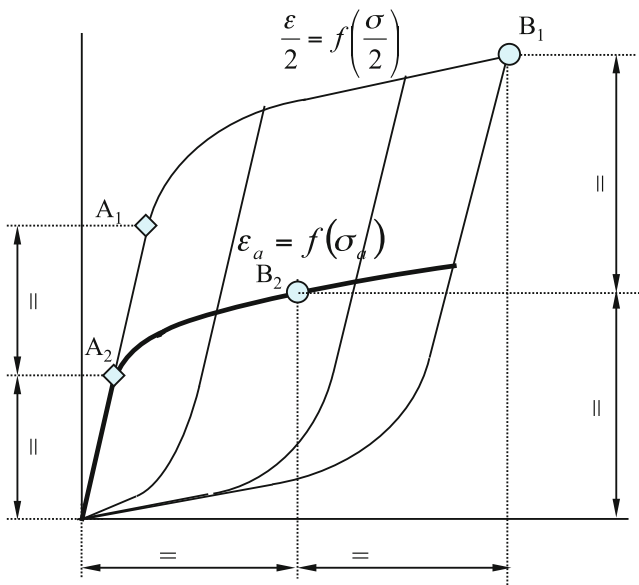


Fig. 8.7 Similarity between the cyclic curve and the branches of the hysteresis loops

them to the origin of the diagram:

$$\frac{\varepsilon - \varepsilon_r}{2} = \left(\frac{\sigma - \sigma_r}{2E} \right) + \left(\frac{\sigma - \sigma_r}{2K'} \right)^{1/n'}$$

$$\frac{\varepsilon_r - \varepsilon}{2} = \left(\frac{\sigma_r - \sigma}{2E} \right) + \left(\frac{\sigma_r - \sigma}{2K'} \right)^{1/n'}$$

σ_r e ε_r are the cuspsid coordinates, i.e. the inversion point of the preceding cycle, taken as the origin.

Both equations can be summarized in one: Eq. 8.2:

$$\frac{\Delta\varepsilon_t}{2} = \frac{\Delta\varepsilon_e}{2} + \frac{\Delta\varepsilon_p}{2} = \left(\frac{\Delta\sigma}{2E} \right) + \left(\frac{\Delta\sigma}{2K'} \right)^{1/n'}. \tag{8.2}$$

8.4.1 Concentration Factors for Variable Loads

Similarly to the static case, at the root of the notch the elastic limit can be exceeded and a small plastic zone is formed. Since the surrounding material behaves elastically, the strain in the plastic zone, as the load increases, tends to be redistributed at the points contiguous to the most loaded ones, with the same mechanism discussed for static loads.

For variable strains, the model of *Neuber* is still appropriate [20] for simulating the local values of stress and strain. The model given for monotonically increasing loads:

$$K_\varepsilon \cdot K_\sigma = K_t^2$$

can be extended to describe the stress and strain that arise at the first monotonic increment of load at the reference point.

For cyclically variable loads, the relation is still valid if it is referred to the new reference system located in the cycle hysteresis cuspsid and the concentration factors are defined in this way:

$$K_\varepsilon = \frac{\Delta\varepsilon}{\Delta e} \quad K_\sigma = \frac{\Delta\sigma}{\Delta S}$$

$$\Delta\varepsilon \cdot \Delta\sigma = K_t^2 \cdot \Delta e \cdot \Delta S \tag{8.3}$$

Equation 8.3, associated with Eq. 8.2, gives the local stress and strain at the root of the notch if the ranges of variation of the nominal stress and strain are known.

With reference to slender bodies, Eq. 8.3 with the coordinates $\Delta\sigma - \Delta\varepsilon$ describes an hyperbola with origin at the reversal points, while the Eq. 8.2 describes the ascending or descending branch of the hysteresis curve. The hyperbola passes through the point with coordinates: $K_t \cdot \Delta e - K_t \cdot \Delta S$, the virtual stress and strain that would occur in the elastic hypothesis, while the common point of both curves gives the real values in elasto-plastic hypothesis.

Exercise 8.1 (*Local Stress and Strain for cyclic deformation*) Suppose, to better highlight the role of variable loads, that the parameters of the material are the same as the static case, Sect. 6.8 in Chap. 6. The material has not suffered, as a result of fatigue, cyclic hardening or softening. Then: $K' = 1,400 \text{ MPa}$, $n' = 0.14$ e $K_t = 3$. The element is subjected to alternate fatigue load which produces a nominal alternate stress:

$$(\Delta S/2) = \pm 360 \text{ MPa}$$

Calculate stress and strain amplitudes at the bottom of the notch under the hypothesis of a stabilized hysteresis loop.

Equations are the following:

$$\Delta \varepsilon \cdot \Delta \sigma = K_t^2 \cdot \Delta e \cdot \Delta S = \frac{K_t^2 \Delta S^2}{E}$$

$$\frac{\Delta \varepsilon}{2} = \frac{\Delta \sigma}{2E} + \left(\frac{\Delta \sigma}{2K'} \right)^{1/n'}$$

Substituting the numerical values $\Delta \varepsilon \cdot \Delta \sigma = 22.22$

$$\frac{\Delta \sigma}{2 \cdot 22.22} - \frac{\Delta \sigma}{2 \cdot 2.1 \cdot 10^5} - \left(\frac{\Delta \sigma}{2 \cdot 1,400} \right)^{1/0.14} = 0$$

from which: $\Delta \sigma = 1,651 \text{ MPa}$ e $\Delta \varepsilon = 1.35 \%$

The hysteresis loop is characterized by the previous data with mean value equal to:

$$\sigma_m = 0 \quad \varepsilon_m = 0.$$

8.5 Fatigue Life Model in Strain Controlled Problems

For the previous reasons, the strain is the dominant factor in Low Cycle Fatigue, and it is more significant to simulate the fatigue life curves in terms of the strain amplitude instead of the stress amplitude.

Curves strain/number of cycles or, better, strain versus the number of reversals in double-logarithmic coordinates, are schematically shown in Fig. 8.8.⁵ The curve obtained by fitting experimental data is described by the following equation:

$$\frac{\Delta \varepsilon}{2} = \frac{\Delta \varepsilon_e}{2} + \frac{\Delta \varepsilon}{2} = \frac{\sigma'_f}{E} (2N)^b + \varepsilon'_f (2N)^c \quad (8.4)$$

⁵ Reversals number (equal to the double number of cycles) distinguishes the semi-cycle with positive strains from the semi-cycle with negative strains and it is more suitable to describe the load history.

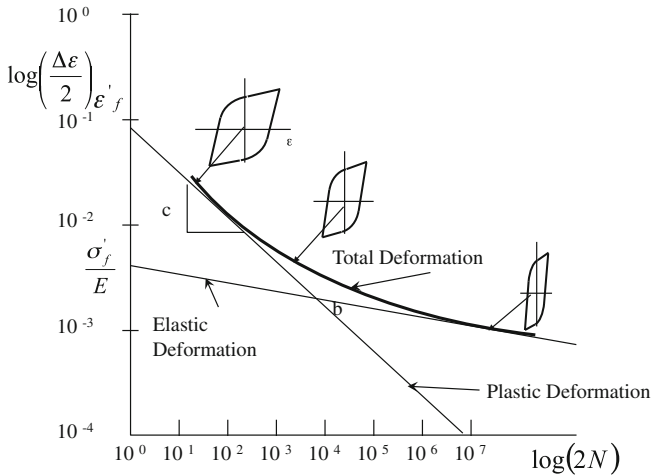


Fig. 8.8 Fatigue life curve with the local strain amplitude compared with the number of load reversals (equal to the double of the load cycles number)

where

$$\begin{cases} \sigma'_f & = \text{fatigue strength coefficient} \\ \epsilon'_f & = \text{fatigue ductility coefficient} \\ c, b & = \text{fatigue strength exponent and fatigue ductility exponent} \\ E & = \text{elasticity modulus} \end{cases}$$

The symbols used in this representation have an historical reason.

In the early years of the theory development, in the absence of direct experimental data, any efforts were made to establish correlations between $\sigma'_f, \epsilon'_f, c$ and b and parameters obtained by static tests.

Today this effort is a non-sense: in any case, methods to simulate the fatigue behavior avoiding fatigue tests must not be encouraged, because they have to be considered scientifically poor and detrimental with respect to the *evidence* of this phenomenological theory and perhaps greatly dangerous in instilling false conclusions. The value of the fatigue tests resides in the fact that they are precisely fatigue tests, i.e. they record the material behavior over the time and clearly demonstrate the cumulate damage phenomenon.

Each term of the sum in Eq. 8.4 in a double-logarithmic diagram is represented by a straight line:

$$\begin{cases} \ln(\sigma'_f/E) & + b \ln(2N) \\ \ln \epsilon'_f & + c \ln(2N) \end{cases}$$

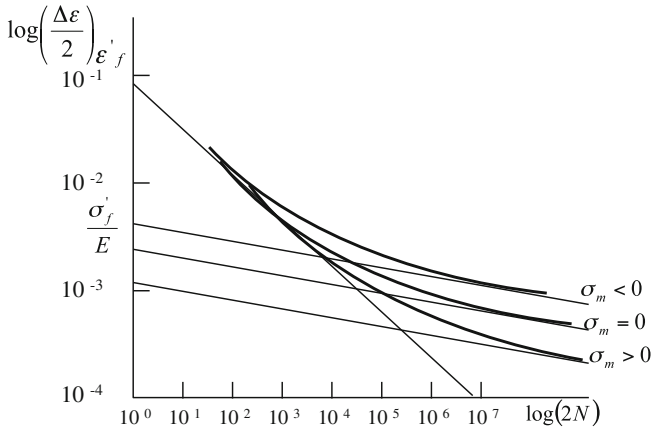


Fig. 8.9 Influence of a mean stress on the strain fatigue curve

The curve $\Delta\varepsilon/2 = f(2N)$ is obtained summing both functions (not the logarithm of the functions!). The fatigue curve is obtained by the logarithm of this sum.

8.5.1 Influence of the Mean Stress

Given a local strain amplitude, the presence of a local average tensile stress in the hysteresis loop reduces the number of load reversals to a first visible crack, and vice-versa for a local mean compression stress. The following model quantifies the extent of this effect:

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma'_f - \sigma_m}{E} (2N)^b + \varepsilon'_f (2N)^c \tag{8.5}$$

It is assumed that the presence of a mean stress changes only the elastic part of the equation, Fig. 8.9 (A mean stress, like other kinds of residual stresses, would be cancelled by plastic deformation). Other models take into account the mean stress influence in different ways [21].

8.6 Model for Variable Amplitudes

If the hysteresis cycles for a variable load are built according to the described models, following the loadings cycles step by step, Fig. 8.10a, the local hysteresis loops are not closed, but some branches of them cut the preceding ones as in the example in the Figure, and the final point 7 does not necessarily coincide with the point 1, Fig. 8.10b.

In the physical reality, the hysteresis larger loop is not affected by smaller intermediate hysteresis loops: when smaller cycles follow a larger incomplete hysteresis cycle if, after such interruptions, the load increases again at the previous level, the

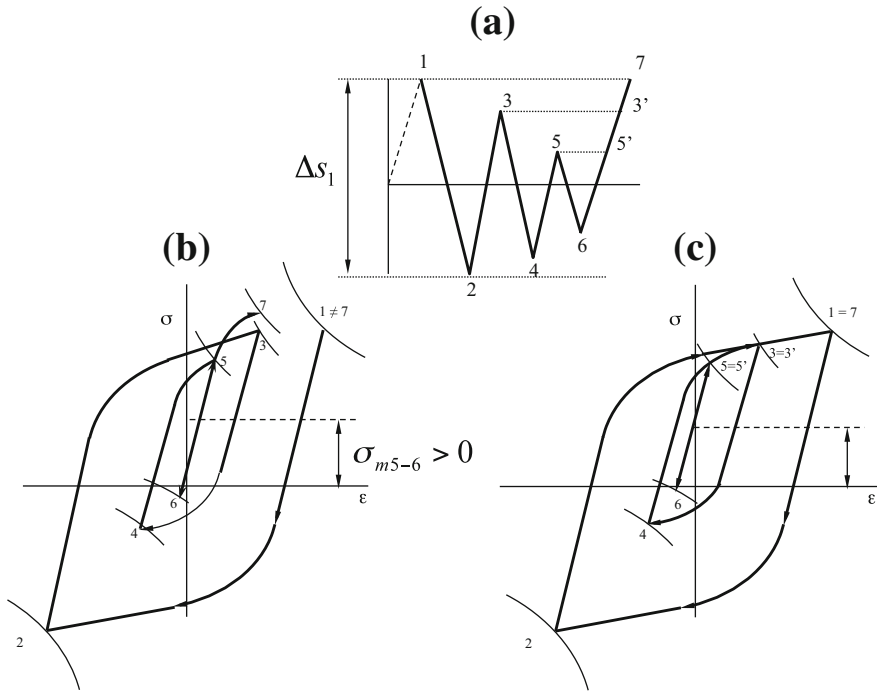


Fig. 8.10 Building of hysteresis cycles by *Ramberg-Osgood* and *Neuber* equations: **a** loads sequence, **b** hysteresis cycles respecting the sequence but not the memory effect of the material and **c** hysteresis cycles respecting the sequence and the memory

material *remembers* the previous history and completes the local stress-strain cycle that had been interrupted, as shown in Fig. 8.10c.

To overcome this discrepancy between reality and model, the steps in the following sections are necessary.

It is remarkable to observe that a local mean stress of the hysteresis loop can arise even in the absence of an applied mean load.

If a series of cycles of constant amplitude is preceded by a cycle of larger amplitude, the subsequent hysteresis cycles assume a local positive or negative mean stress depending on whether the first starting branch of the cycle is increasing in a positive or in negative direction.

In other words, the actual sequence of load cycles is responsible for the onset of the states of positive or negative residual stress, thus of the fatigue life of the component itself. Life can be shortened or lengthened compared to the case of the absence of the first load cycle.

A mean non-zero local stress can thus occur in the absence of the mean value of the load and the sign depends upon the sequence of load cycles.

8.6.1 Cycle Counting Method

For High Cycle Fatigue estimation it was necessary to develop methods of counting the load cycles for reducing and synthesizing the loads history in stress spectra for the amplitudes and mean values.

For Low Cycle Fatigue the counting procedures also have the task of establishing a procedure for building the local hysteresis cycles, respecting the real loads sequence, since the overcoming of the plastic limits is an irreversible action that needs the temporal load sequence to be respected.

The real load sequence must be reduced to an equivalent sequence of *closed* load cycles. This is the first step in order to generate local closed hysteresis loops, through which a damage evaluation is possible. For this purpose a counting method due to Matsuishi and Endo [22] named *Rainflow*, is able to identify closed load cycles equivalent to real load sequence.⁶

If the method is applied to the sequence of loads shown in Fig. 8.10a, three equivalent closed loading cycles are identified that generate the following closed hysteresis loops (1–2–7) (3–4–3') (5–6–5').

The method starts by identifying as first the larger load cycle, cutting out with horizontal lines cycles of smaller amplitudes, and subsequently identifying the smaller cycles in descending order, using the same rule. This method of counting, even if it reduces the actual series to a series of closed cycles, keeps the actual sequence of hysteresis loops, respecting the memory of the material.⁷

In this way, the different hysteresis loops are obtained *closed* and in proper sequence, *attached* one after the other, at the points identified by the intersection between the ascending branch of the hysteresis cycle and the *Neuber* hyperbola, that have the same origin as the hysteresis branch. By means of this artifice the hysteresis loops take the appearance of Fig. 8.10c.

8.6.2 Damage Per Cycle and Accumulated Damage

The way to assess the local damage due to a cycle is the following: for each i th cycle the amplitude ε_i and the mean stress σ_{mi} are determined and can enter the mathematical model of Eq. 8.5. Solving this equation, or utilizing the corresponding fatigue curve of Fig. 8.9, gives the reversals to failure $2N_i$ for the i th cycle.

Cycle i th is responsible for a life reduction equal to $1/N_i$. The damage due to n_i repetitions of the same cycle is then n_i/N_i . If a sequence of hysteresis loops (constituting a group of cycles that are periodically repeated), is formed by k cycles,

⁶ The concept of equivalence is related to an equivalent damage produced by the sequence of the hysteresis loops generated by real and by the virtual loads.

⁷ Properly only if the load cycles are followed in descending order.

the damage of a group of k cycles can then be estimated by a linear accumulation law as:

$$\sum_{i=1}^k \frac{n_i}{N_i}$$

and the damage for the repetition of this sequence for a number m of times is:

$$D = m \cdot \sum_{i=1}^k \frac{n_i}{N_i} \quad (8.6)$$

This is an application of the so-called *The Palmgren-Miner's rule*, that, like the *Rain-Flow Counting Method*, is a general rule in Fatigue, not limited only to Low Cycle Fatigue, because this rule can be used for stress as well for strain-controlled phenomenon.

The *Palmgren-Miner's rule* [23], assumes a linear damage hypothesis. According to this theoretical assumption, failure would occur for $D = 1$. This can be a good estimation in a design phase. Nevertheless, experimental evidence shows that failure occurs when D is larger than 1, with values up to 2.2.

A doubt can arise on its capability to predict effects in a phenomenon largely influenced by the load history, i.e. by the order in which the reversals occur.

In High Cycle Fatigue, where stresses remain always in elastic field, a linear combination of cycles damage can be accepted.

In Low Cycle Fatigue the described counting method makes it possible to build the local hysteresis loops in the real sequence in which they are generated, respecting the loads sequence. After this treatment, a linear damage accumulation law is used.

The described equivalent sequence of closed load cycles takes into account the effect of an overload that generates a residual stress state that is the real factor responsible for the sequence effect: in fact, a compressive residual stress visualized as negative mean stress, may delay the crack growth and vice-versa for a positive mean stress. The Palmgren-Miner rule can fail to recognize the probabilistic nature of fatigue. For this reason, in a Chap. 11 the use of design curves will be adjusted to account for scatter of experimental data.

8.6.3 Application of Neuber Hypothesis Without Concentration Factors

When concentration factors are not known, the *Neuber* equation: Eq. 8.7.

$$\Delta \varepsilon \cdot \Delta \sigma = K_t^2 \cdot \Delta e \cdot \Delta S \quad (8.7)$$

can be modified in the following way that avoids the introduction of concentration factors:

$$\Delta\varepsilon \cdot \Delta\sigma = \Delta\varepsilon_{elastic} \cdot \Delta\sigma_{elastic}$$

According to this observation, the area of the rectangle built with the coordinates of the real point is identical to the area built with the coordinates of the virtual point that represents the perfectly elastic behavior. These can be numerically computed by an elastic model.

Since the equation of the ascending or descending branch of the hysteresis loop is known for that material, it is therefore possible to determine the local hysteresis cycle. Nevertheless, the calculation is approximate because it implies the simplified *Neuber* model.

Exercise 8.2 (*First monotonic load followed by fatigue cycles that produces stabilized hysteresis cycles*) With the data of Exercise 8.1 determine the local stabilized hysteresis cycle, assuming a new (not previously loaded) material with a nominal stress equal to:

$$S_1 = +460 \text{ MPa} \quad S_2 = -260 \text{ MPa}$$

(a) It is necessary to distinguish the tension side from the compression side. The loading phase from zero to S coincides with the monotonic curve. Equations to describe the load increment from 0 to S_1 are the following:

$$\begin{aligned} \varepsilon \cdot \sigma &= K t^2 \cdot e \cdot S = K t^2 (S^2/E) \\ \varepsilon &= (\sigma/E) + (\sigma/K)^{1/n} \end{aligned}$$

Solving both equations gives the coordinates σ_{r1} and ε_{r1} of the point.

(b) In the descending phase from S_1 to S_2 the equations are:

$$\begin{aligned} (\varepsilon_{r1} - \varepsilon)(\sigma_{r1} - \sigma) &= K_t^2 \cdot \frac{(S_1 - S_2)^2}{E} \\ \frac{\varepsilon_{r1} - \varepsilon}{2} &= \frac{\sigma_{r1} - \sigma}{2E} + \left(\frac{\sigma_{r1} - \sigma}{2K'} \right)^{1/n'} \end{aligned}$$

The solutions of both equations gives the coordinates σ_{r2} ε_{r1} .

The hysteresis loop closes with the load passing from S_2 to S_1 , because the point is reached with the coordinates σ_{r1} , ε_{r1} . Substituting these data in the previous equations gives: σ_{r1} , ε_{r1} and σ_{r2} , ε_{r2} .

(c) The loop parameters are:

$$\Delta\sigma = \sigma_{r1} - \sigma_{r2}$$

$$\Delta\varepsilon = \varepsilon_{r1} - \varepsilon_{r2}$$

$$\sigma_m = \frac{\sigma_{r1} + \sigma_{r2}}{2}$$

$$\varepsilon_m = \frac{\varepsilon_{r1} + \varepsilon_{r2}}{2}$$

The preceding hysteresis cycle *pulls* the hysteresis cycle that follows it. This explains the role of the load history, i.e. the irreversible effect of load cycles.

As a consequence, the mean values of σ and ε of a hysteresis loop are not directly related with the mean value of the nominal stress, since they also depend on the hysteresis cycles that precede. That confirms how mean stress and strain can be generated at the bottom of a notch also in the case of variable load amplitudes with mean values equal to zero.

The calculation of the mean local stress and strain is important because they influence the fatigue life of the component.

8.7 Case Study of a Damage Evaluation

To overcome the approximation of all simulation models, a direct deformation measurement, instrumenting and recording the local strain, cycle after cycle, at the points where nucleation of first defects is expected, would be necessary.

Strain gages with a size properly selected in comparison to the deformation gradient, can be usefully employed. It is not necessary to use strain gages particularly resistant to fatigue [24], because the strain can be recorded only for a sequence of hysteresis loops which is repeated periodically, as usually happens in machines components.⁸

In some cases, the application of experimental analysis may be the only method for solving complex simulation problems together with a numerical model for the counting of the hysteresis loops and for the damage assessment. Component life is thus evaluated by a hybrid numerical-experimental approach as follows [25]:

1. The load sequence (with load variable amplitudes) is applied to the component. Often this phase is an *in-field* measurement on real components with real loading conditions.
2. A diagram of the local deformation versus time is recorded by strain gages placed at the reference point. If the deformations in both principal directions vary proportionally or at least in phase to each other, the concept of the equivalent strain is applied.⁹
3. The recorded strain sequence at the reference point is applied to a smooth specimen of the same material.

⁸ Strain gages provided for large deformations have a very limited fatigue resistance and cannot be used for long load sequences.

⁹ Otherwise, criteria of multiaxial fatigue [2, 26, 27], must be used.

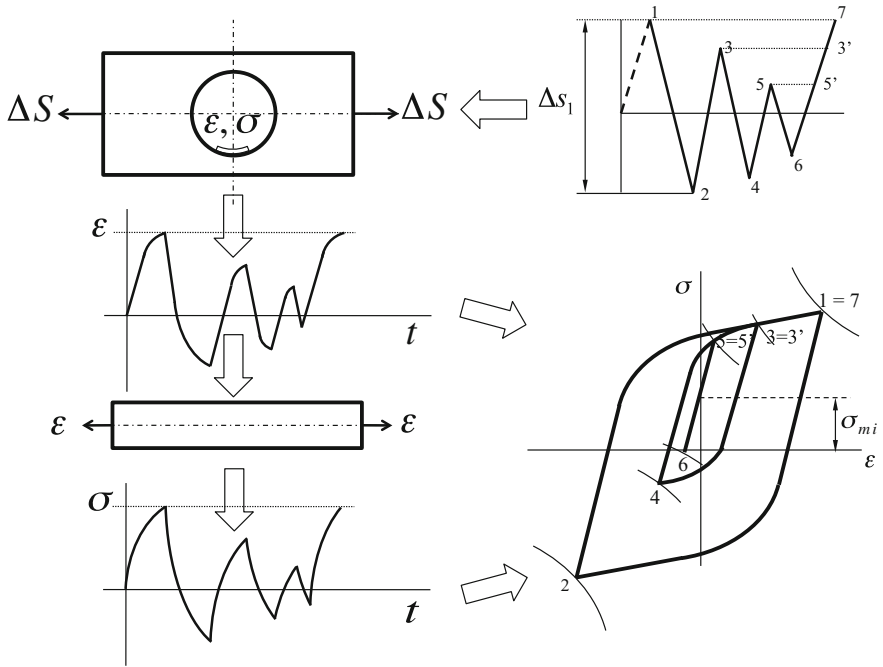


Fig. 8.11 Experimental determination of the sequence of hysteresis loops in a geometrical discontinuity

4. The corresponding stress history on this specimen is recorded and stored in a computer memory together with the strain history.
5. The sequence of the local hysteresis cycles is derived from strain and stress histories at the reference point, Fig. 8.11.
6. Local damage is estimated for each hysteresis cycle and for all cycles.

Figure 8.12 shows an example for an earth-moving machine where, in a discontinuity of a complex component, fatigue cracks were revealed in previous experiments.

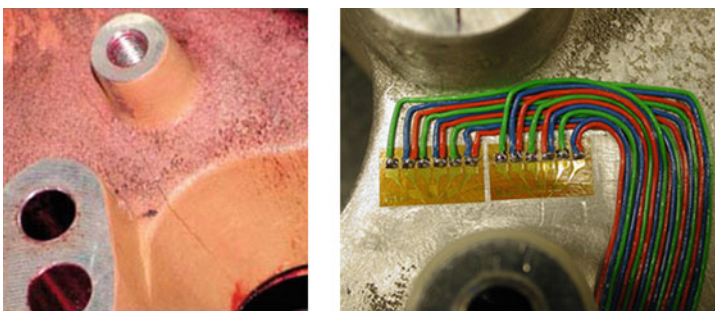


Fig. 8.12 Example of strain gage measurements at a reference point

A new component is mounted in place of the fractured one and instrumented by strain gages set out in the direction perpendicular to the expected cracks. The earth-moving machine was operated with a sequence of movements, in such a way that the same load history as the replaced component was applied to the instrumented one. In this way a reliable damage assessment is done.

Similar approach for the estimation of the fatigue life in turbo-machinery on experimental base, is applied in the following papers [28, 29].

8.8 Torsional Low Cycle Fatigue

A general relationship between stresses and strains in the case of multi-axial states of stress in the simplest case of proportional (or in-phase) loading condition, is introduced by the concept of octahedral tangential stress and strain [30–33]:

$$\tau_{oct} = f(\gamma_{oct}) \quad (8.8)$$

or its inverse:

$$\gamma_{oct} = g(\tau_{oct}) \quad (8.9)$$

where:

$$\tau_{oct} = \frac{1}{3} [(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2]^{1/2} \quad (8.10)$$

and:

$$\gamma_{oct} = \frac{2}{3} [(\varepsilon_1 - \varepsilon_2)^2 + (\varepsilon_2 - \varepsilon_3)^2 + (\varepsilon_3 - \varepsilon_1)^2]^{1/2} \quad (8.11)$$

The following relationship is valid in the elastic field :

$$\tau_{oct} = G \cdot \gamma_{oct} \quad (8.12)$$

Vice-versa, to build a theory on hardening the following assumptions are required:

- Small components of the strain state.
- Isotropic material yield.
- Invariance of the material volume when it is deformed beyond the elastic limit ($\nu = 0.5$).
- Negligible influence on the permanent deformation of mean stress $\sigma_m = (\sigma_1 + \sigma_2 + \sigma_3)/3$ (normal stress on the octahedral planes).

With these limitations, Eq. 8.9 can be used to establish comparisons between the results of tests in conditions of uniaxial and multi-axial loadings, in particular between uniaxial and torsional loading [34].

For a uniaxial load the relationship between octahedral and uniaxial parameters is:

$$\tau_{oct} = \frac{\sqrt{2}}{3} \sigma \quad (8.13)$$

For the shear strain expression, two different equations exist, depending on the value of the Poisson's ratio¹⁰:

$$\begin{cases} \gamma_{oct} = \sqrt{2}\varepsilon & \text{for } \nu = 0.5 \\ \gamma_{oct} = \frac{2}{3}\sqrt{2}(1+\nu)\varepsilon & \text{for } \nu \neq 0.5 \end{cases} \quad (8.16)$$

and for torsional loads:

$$\begin{cases} \tau_{oct} = \sqrt{\frac{2}{3}}\tau \\ \gamma_{oct} = \sqrt{\frac{2}{3}}\gamma \end{cases} \quad (8.17)$$

where τ is the maximum stress tangential to the shaft surface.

8.8.1 Ramberg-Osgood Model for Torsional Loads

Material hardening for cyclic load for multi-axial proportional stress state is described by the following relationship:

$$\gamma_{oct} = \gamma_e + \gamma_p = \frac{\tau_{oct}}{G} + \left(\frac{\tau_{oct}}{C'}\right)^{1/n'} \quad (8.18)$$

¹⁰ In order to overcome this difficulty a variable expression of ν is introduced that must be determined by the cyclic stress-strain curve [35, 36]:

$$\bar{\nu} = 0.5 - (0.5 - \nu_e) \frac{E_s}{E} \quad (8.14)$$

where:

$$\begin{cases} \bar{\nu} & = \text{variable Poisson coefficient} \\ E & = \text{elasticity modulus} \\ E_s & = \text{secant modulus on the cyclic curve} \\ \nu_e & = \text{Poisson's coefficient in the elastic field} \end{cases}$$

Secant modulus E_s is obtained from the cyclic curve utilizing the effective strain $\bar{\varepsilon}$:

$$\gamma_{oct} = \frac{2}{3}\sqrt{2}(1+\bar{\nu}) \cdot \bar{\varepsilon}$$

that for Eq.8.11 gives:

$$\bar{\varepsilon} = \frac{1}{\sqrt{2}(1+\bar{\nu})} [(\varepsilon_1 - \varepsilon_2)^2 + (\varepsilon_2 - \varepsilon_3)^2 + (\varepsilon_3 - \varepsilon_1)^2]^{\frac{1}{2}}. \quad (8.15)$$

where:

$$\left\{ \begin{array}{l} \tau_{oct} \quad = \text{amplitude of tangential octahedral stress} \\ \gamma, \gamma_e, \gamma_p \quad = \text{amplitude of total, elastic and plastic, tangential octahedral stress.} \\ G \quad = \text{tangential elastic modulus.} \\ C' \quad = \text{cyclic strength coefficient.} \\ n' \quad = \text{cyclic strain hardening exponent.} \end{array} \right.$$

1. For axial load [3], Eq. 8.18 is reduced to the following, substituting Eq. 8.13 and the first of Eq. 8.16 in Eq. 8.18:

$$\varepsilon = \frac{\sigma}{E} + \left(\frac{\sigma}{K'} \right)^{1/n'} \tag{8.19}$$

where the cyclic strength coefficient for axial case K' holds:

$$3 \cdot 2^{(n'-1)/2} \cdot C' \tag{8.20}$$

2. For torsional alternate load, substituting Eq. 8.17 in Eq. 8.18, we obtain:

$$\gamma = \frac{\tau}{G} + \left(\frac{\tau}{H'} \right)^{1/n'} \tag{8.21}$$

with:

$$H' = 2^{(n'-1)/2} \cdot 3^{(1-n')/2} \cdot C' \tag{8.22}$$

where:

$$\left\{ \begin{array}{l} \tau \quad = \text{amplitude of shear stress} \\ G \quad = E/2(1 + \nu) \text{ shear elastic modulus} \\ H' \quad = \text{cyclic strength coefficient in torsion} \end{array} \right.$$

Exponent n' remains unchanged for axial and torsional stress.

Cyclic torsional strength coefficient H' can be obtained from the cyclic strength coefficient for axial load, comparing Eq. 8.21 with Eq. 8.19:

$$H' = K' \left(\frac{1}{3} \right)^{(n'+1)/2} \tag{8.23}$$

We can summarize the points observing that, if Eqs. 8.8 or 8.9 are appropriate, i.e. if the state of multi-axial stresses and strains can be described by octahedral shear stress and shear strain, it is possible to experiment under uni-axial load conditions and to obtain the parameters of the cyclic curve in torsion. The Ramberg-Osgood parameters for torsional loadings H' are obtained by Eq. 8.23 with Eq. 8.21 instead of Eq. 8.19.

8.8.2 Neuber Model for Torsional Loads

Neuber formula [37], becomes:

$$\Delta\tau \cdot \Delta\gamma = K_t^2 \cdot \Delta T \cdot \Delta\Gamma \quad (8.24)$$

where:

$$\begin{cases} \Delta\tau & = \text{local shear stress range} \\ \Delta\gamma & = \text{local shear strain range} \\ \Delta T & = \text{nominal shear stress range} \\ \Delta\Gamma & = \text{nominal shear strain range } (= \Delta T / G) \end{cases}$$

Going from Eqs. 8.7–8.24 needs the following substitutions:

γ instead of ε
 τ instead of σ
 G instead of E .

8.8.3 Fatigue Curve for Torsional Loads

The simulation of the uni-axial fatigue curve is:

$$\frac{\Delta\varepsilon}{2} = \varepsilon_a = \frac{\sigma'_f}{E} (2N_f)^b + \varepsilon'_f (2N_f)^c \quad (8.25)$$

Similarly, in the case of torsional fatigue the points that correlate the amplitude of shear strain to the number of reversals to failure can be interpolated by a relationship of the same type.

$$\frac{\Delta\gamma}{2} = \gamma_a = A (2N_f)^\alpha + B (2N_f)^\beta \quad (8.26)$$

Even these two equations may be correlated by the theory of octahedral tangential shear [4, 38, 39].

Equating the octahedral torsional shear strains to the axial case, the link between the constants σ'_f/E , ε'_f , b and c and A , B , α e β is obtained.

1. In plastic field relationships Eqs. 8.16 and 8.17, are still valid; since the octahedral shear strains γ_{oct} must be equal, the following holds:

$$\sqrt{2} \varepsilon = \sqrt{\frac{2}{3}} \gamma$$

from which:

$$\varepsilon = \frac{\gamma}{\sqrt{3}} \quad (8.27)$$

It must be thus:

$$\varepsilon'_f (2N_f)^c = \frac{B}{\sqrt{3}} (2N_f)^\beta$$

then:

$$\begin{cases} c = \beta \\ \varepsilon'_f = B/\sqrt{3} \end{cases} \quad (8.28)$$

2. For the elastic contribution, remembering that the condition of volume conservation (assumed for the plastic part of the deformation) is not valid, equating the octahedral shear strain gives, for the second relationship of Eq. 8.16:

$$\frac{2}{3} \sqrt{2} (1 + \nu) \varepsilon_1 = \sqrt{\frac{2}{3}} \gamma$$

then:

$$\varepsilon_1 = \frac{\sqrt{3}}{2(1 + \nu)} \gamma \quad (8.29)$$

Equating the elastic parts of Eqs. 8.25 and 8.26, we have:

$$\frac{\sigma'_f}{E} (2N_f)^b = \frac{\sqrt{3}}{2(1 + \nu)} A (2N_f)^\alpha$$

and finally:

$$\begin{cases} b = \alpha \\ A = \frac{2}{\sqrt{3}} (1 + \nu) \frac{\sigma'_f}{E} \end{cases} \quad (8.30)$$

With this assumption the fatigue curves derived under the conditions of uniaxial stress can be interpreted as fatigue curves for torsional by this equation:

$$\gamma_a = A (2N_f)^\alpha + B (2N_f)^\beta \quad (8.31)$$

with:

$$\begin{aligned} A &= \frac{2}{\sqrt{3}} (1 + \nu) \frac{\sigma'_f}{E} & \alpha &= b \\ B &= \sqrt{3} \cdot \varepsilon'_f & \beta &= c \end{aligned}$$

The algorithms developed for the uniaxial case can still be used for the torsional case, replacing ε with γ and σ'_f/E , ε'_f , b and c with the new constants A , B , α and β given by Eq. 8.31.

8.8.4 Effect of Mean Shear Stress

The influence of the mean stress in the case of torsion tests must be assessed with a different criterion to the uniaxial case, since now it is not possible to distinguish the damage on the basis of the stress sign.

A new damage function is introduced [38], of this type:

$$P_{swt} = \sqrt{|\tau_{max}| \cdot \gamma_a \cdot G} = \sqrt{|\tau_m + \tau_a| \cdot \gamma_a \cdot G} \quad (8.32)$$

The double-logarithmic diagram of P_{swt} versus the reversals to failure, reduces the data obtained from different tests (with several values of the ratio $R = \sigma_{min}/\sigma_{max}$) to a single curve that still has the characteristic shape of a fatigue curves.

The *Manson-Coffin* equation for zero mean stress and its elastic part is:

$$\begin{cases} \gamma_a = A (2N_f)^\alpha + B (2N_f)^\beta \\ \tau_\alpha = A \cdot G \cdot (2N_f)^\delta \end{cases} \quad (8.33)$$

If P_{swt} is expressed as a function of the maximum stress τ_{max} and of the amplitude of the equivalent stress that, for zero mean stress, produces failure for the same reversals number, the following equivalence can be written:

$$\sqrt{\tau_{max} \cdot \gamma_a \cdot G} = \sqrt{\tau_{aeq} \cdot \gamma_a \cdot G}$$

valid for a given number of reversals to failure. Since τ_{aeq} is:

$$\tau_{aeq} = A \cdot G \cdot (2N_f)^\delta$$

then:

$$\tau_{max} \gamma_a = AG (2N_f)^\alpha [A (2N_f)^\alpha + B (2N_f)^\beta]$$

that can be reduced to the form:

$$\tau_{max} \cdot \gamma_a = A^2 \cdot G \cdot (2N_f)^{2\alpha} + ABG (2N_f)^{\alpha+\beta} \quad (8.34)$$

Equation 8.32 becomes:

$$P_{swt} = \sqrt{\tau_{max} \cdot \gamma_a \cdot G} = [(AG)^2 (2N_f)^{2\alpha} + ABG^2 (2N_f)^{\alpha+\beta}]^{1/2} \quad (8.35)$$

The damage function P_{swt} has the dimension of a stress. It can be normalized dividing P by $\tau = \sigma/\sqrt{3}$ [38]:

$$P_{swt}^* = \frac{P_{swt}}{\tau}$$

(a dimensionless parameter).

8.9 Case Study of Notched Shafts in Torsion

The present case study [40–42], shows the simulation of the low cycle fatigue life for a shaft with circumferential notches, as shown in Fig. 8.13.

In order to create a simulation model for this case, the following steps are developed, having defined the kinds of loads on the prototype (constant amplitude or variable amplitude loads and the relative load sequence):

- Utilizing special equipments for material characterization in shear strain controlled mode. This could require the development of special torsion machines with capacity appropriate to the prototype dimensions.
- Determining the cyclic curve parameters for torsional loads.
- Determining the *Manson-Coffin* curve parameters for torsional loads.
- Performing the torsional tests.

8.9.1 Special Equipment for Material Characterization in Shear Strain-Controlled Mode

Figure 8.14 shows a device developed for transforming the tensile load-compression operation of a universal testing machine, into an alternating torsion machine at low-frequency. The same concept, designed with high stiffness members, is developed for high frequency tests, like the device shown in Chap. 7.

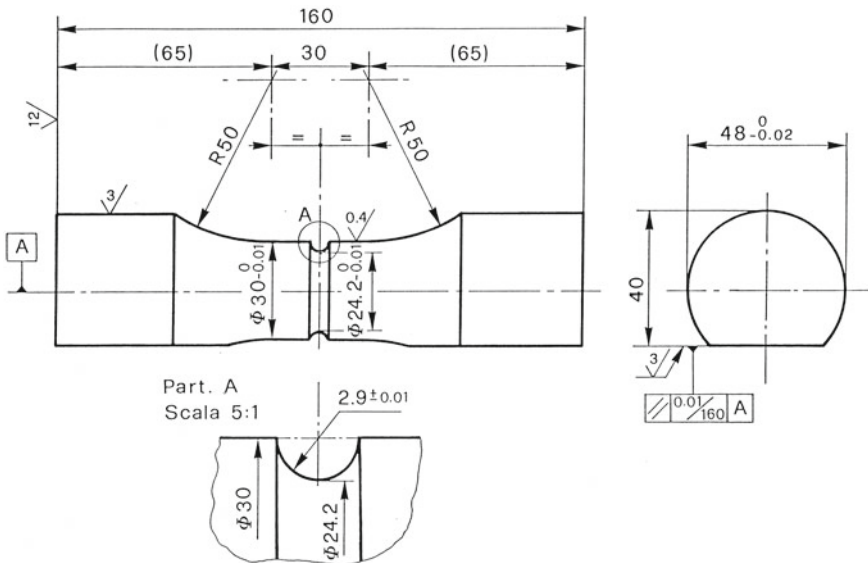


Fig. 8.13 Notched cylindrical specimen for torsion tests

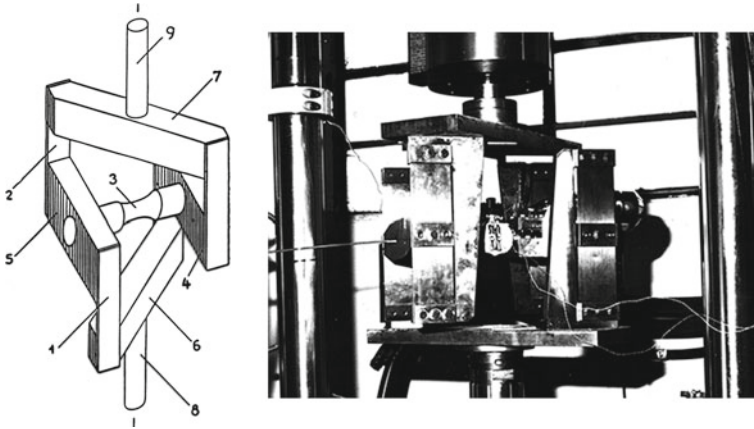


Fig. 8.14 Self-manufactured testing apparatus for torsional low cycle fatigue tests (developed by A. Maggiore Private communication)

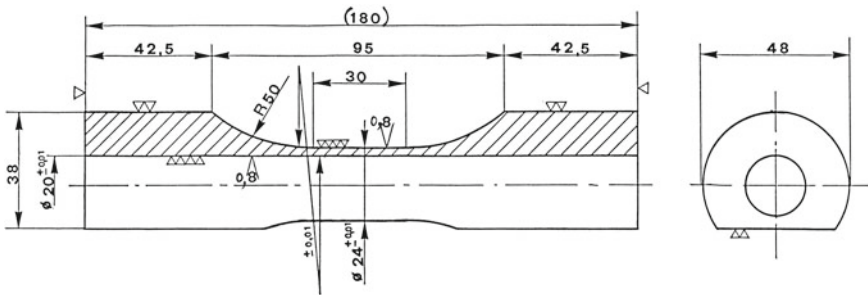


Fig. 8.15 The hollow cylindrical specimen used for material characterization in low cycle fatigue

Fatigue parameters of this material are obtained from hollow cylindrical specimens shown in Fig. 8.15 by shear strain-controlled test. The investigated material is a Low Alloy Steel for Turbine shafts, Table 8.1.

8.9.2 Determination of Cyclic Curve and Fatigue Curve Parameters

The torsional cyclic curve obtained from the hollow specimens is shown in Fig. 8.16, Eq. 8.21. The dotted line is the octahedral prevision. The fatigue curve is obtained from the same specimens and is shown in Fig. 8.17. The parameters of both models are reported in Table 8.2.

Table 8.1 Data on a steel tested for low cycle fatigue

C	Si	Mn	P	S	Cr	Mo	Ni	V	Cu	Al	Rs	Rm	A %	z %	Charpy V (J)	FATT (°C)
0.28 %	0.07	0.23	0.008	0.004	1.63	0.42	3.59	0.09	0.06	0.009	681	807	20.1	64.2	135	-25

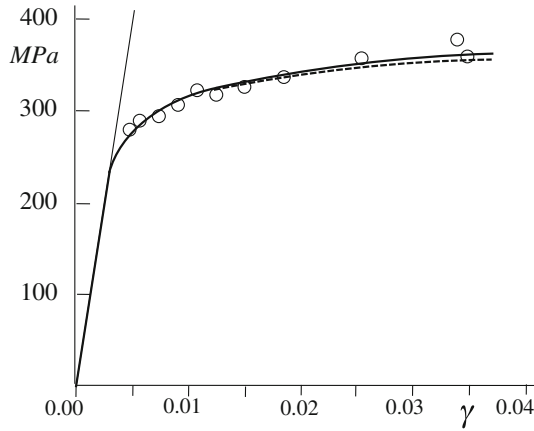


Fig. 8.16 Cyclic curve for test in torsion of hollow cylindrical specimens. Mat. data: $G = 77,990$ MPa $H' = 504.5$ MPa $n' = 0.092$. Circular marks experimental data for hollow specimens. The dotted line is the octahedral prevision

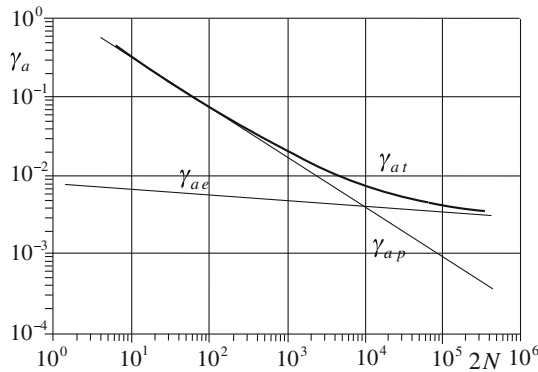


Fig. 8.17 Shear strain fatigue curve for torsional test of hollow cylindrical specimens, Eq. 8.31: $A = 0.0067$ $\alpha = -0.058$ $B = 1.39$ $\beta = -0.63$

Table 8.2 Torsional low cycle fatigue parameters

E (MPa)	G (MPa)	H' (MPa)	n'	A	B	α	β
202,774	77,990	504.58	0.0919	6.7×10^{-3}	1.39	-0.058	-0.63

8.9.3 Special Equipments for Torsional Tests on Shafts

Figure 8.18 shows a special testing machine built for testing low cycle torsion fatigue on shafts. A crank mechanism applies an alternating torque up to ± 25 kN m to the shafts specimens mounted in the machine.

Fig. 8.18 Self-manufactured testing apparatus built for alternate torsion testing with loading capacity up to $\pm 25 \text{ kN m}$



The cell, Fig. 8.19, is made with a tubular body (on the right side), instrumented by four electrical strain gages placed at 45° with respect to the axis, with the full-bridge scheme described in the Chap. 2, Fig. 2.17, by the equation:

$$\frac{\Delta e}{e_0} = \frac{K}{4} \cdot (\varepsilon_1 + \varepsilon_1 + \varepsilon_1 + \varepsilon_1) = \frac{4K}{4} \cdot \varepsilon_1 = K \varepsilon_1 \quad (8.36)$$

8.9.4 Torsional Tests with Assigned Loads Sequence

The first tests are planned with constant amplitudes torsional loads. The sequences derived by in-field measurements are successively applied.

It is acceptable practice to simplify the in-field measurements, condensing the peak loads and neglecting the smallest load cycles that have low influence on the damage. After this first screening, load sequences with descending peaks have more interest than sequences in which the largest peak is preceded by cycles of lower amplitude.

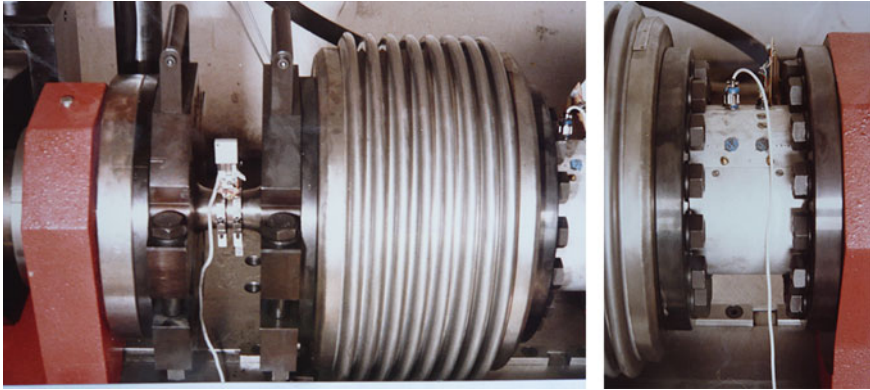


Fig. 8.19 On the *left*, a control device of the torsional deformation with a below for absorbing spurious bending and, on the *right*, the loading cell

Table 8.3 Load sequences for low cycle fatigue tests

Sequence	H	I	L	M	Peak amplitude moment (Nm)	Ratio	Nominal Amplit. (MPa)
I	1	4	8	5	1,502	1	423
II	12	50	100	62	1,197	0.8	342
III	112	450	900	562	900	0.6	279
IV	625	2,500	5,000	62	3,125	0.5	252
Life (cycles)	6,000	3,000	5,750	4,754			

In this case study the sequences of constant and of variable amplitudes reported in Table 8.3, have been considered.

An example of the comparisons between simulated and measured cycles numbers is shown in Fig. 8.20, based on the following models:

1. Estimation of the equation of the branches of hysteresis loops obtained from the cyclic curve as previously presented.
2. Use of the *Neuber* equation for each local cycle.
3. Use of the *Manson-Coffin* equation taking into account the local mean stress of each hysteresis loop.
4. Use of the *Rain Flow* counting method.

8.9.5 Appearance of the Cracks Directions in Torsional Low Cycle Fatigue

Figures 8.21 and 8.22 show different kinds of fracture appearances of cylindrical hollow and solid specimens, loaded by strain-controlled torsional low cycle fatigue.

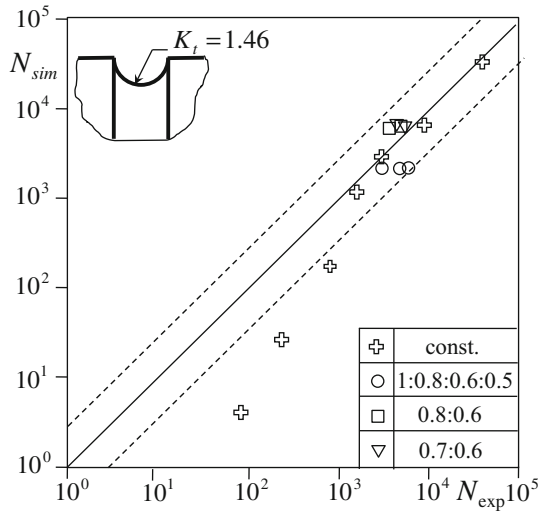


Fig. 8.20 Comparison between simulated and experimental fatigue life (number of cycles to first visible crack) for specimens of Fig. 8.13

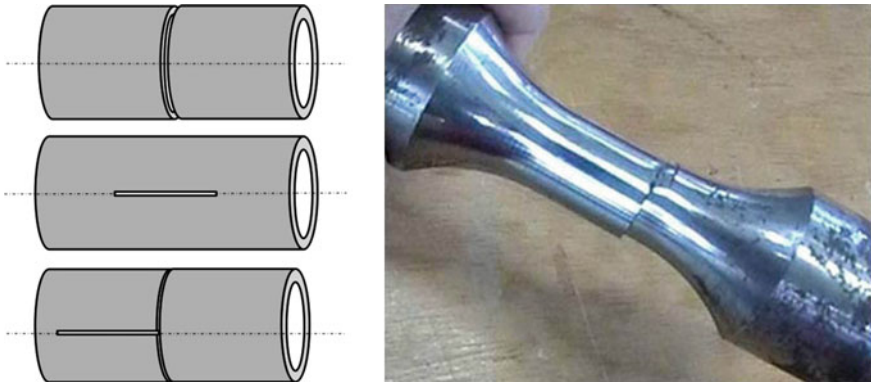


Fig. 8.21 Cracks directions in hollow specimens of 25 mm, loaded by shear-controlled torsional low cycle fatigue ($R = -1$)

In hollow specimens of small thickness, with a quite uniform distribution of shear stress in the cross-section, cracks lines follow the directions of τ_{max} that occurs either at the circumferential direction or along the specimen axis.

In the case of solid specimens, (with triangular distribution of shear stress along the cross-section), the fracture is developed along the cross-section, differently from the typical fracture shapes for high cycle fatigue, see Chap. 7.

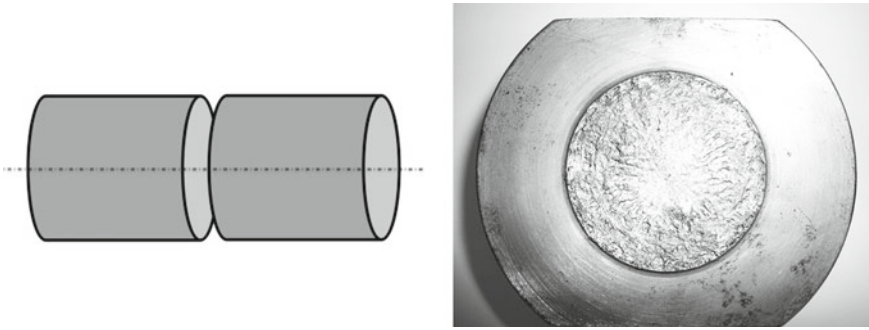


Fig. 8.22 Crack directions in solid specimens of 50 mm loaded by shear-controlled torsional low cycle fatigue ($R = -1$)

8.10 Preventing Loads Misalignments

Many types of loading conditions are theoretically possible for material characterization, but the adoption of an axial load is the most common approach for static and cyclic tests, see standards [43–46].

When performing a test, where a specimen is loaded by a longitudinal force, it is important to prevent or drastically reduce misalignments, which could induce undesired bending or even buckling, when a compressive component is present. Additional bending implies a severe reduction of resistance [47, 48] and the falsification of the test results.

Besides, when an experimental research is developed in cooperation with several laboratories (with tests performed on different machines), a problem arises when the results are compared and collated together, due to the scatter of data, mainly because of uncertainties in measurement. In Refs. [49–51] it is remarked that one of the main causes for these uncertainties is the misalignment of the test piece with respect to the axial load.

Misalignments may be due to a poor alignment in the loading train or to a bad alignment between the fixed and the moving cross-heads of the machine.

- The first type of misalignment is usually greatly reduced by adopting very rigid frames and stiff coupling tolerances between the actuator rod and its bearings.
- The second type of misalignment is often related to a lateral or an angular displacement between the machine grips.

Some Standards, such as [43, 44], recommend a careful check of misalignment, usually achieved by an instrumented specimen, but provide very little information about the technical solutions to be utilized to fulfill an acceptable alignment.

8.10.1 Functional Analysis of the Fixtures

Functional analysis [52] may be suitable to split the global function of the fixture into sub-functions.

All the conceptual solutions for sub-functions F_1 , F_2 and F_3 are summarized in Fig. 8.23 in the form of a *Morphological matrix* [52, 53]. Three main sub-functions, referred to as F_1 , F_2 , and F_3 , can be detected.

The first one, F_1 , refers to complete constraining at one specimen end, whereas the second one, F_2 , is the most difficult to achieve, since it involves constraining at the other end, together with the possibility of adjustment either at the beginning of the test or as a self-alignment during the test.

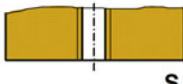
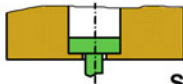

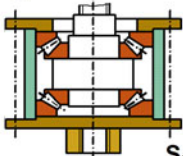
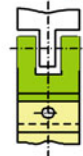

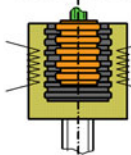

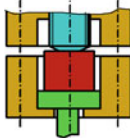
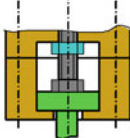
CONCEPTUAL SOLUTIONS			
F_1	<p>Threaded hole</p>  <p>S_{11}</p>	<p>Shape coupling</p>  <p>S_{12}</p>	
F_2	<p>Sphere coupling</p>  <p>S_{21}</p>	<p>Duplex orient. bearing</p>  <p>S_{22}</p>	<p>Universal joint</p>  <p>S_{23}</p>
	<p>Bending elements</p>  <p>S_{24}</p>	<p>Wood's metal</p>  <p>S_{25}</p>	<p>Thrust rolling bearing</p>  <p>S_{26}</p>
F_3	<p>Calibrated disk</p>  <p>S_{31}</p>	<p>Screw tightening</p>  <p>S_{32}</p>	

Fig. 8.23 Morphological matrix [52] containing the three sub-functions and the available conceptual solutions

Sub-function F_3 guarantees that no clearance occurs at force sign inversion, through a slight pre-compression at the specimen ends to prevent any clearance between the transverse faces at the ends and the loading device.

1. The first sub-function (F_1) can be easily achieved by threaded or button-head connections at specimen ends, as suggested for instance by [43, 44].
2. Regarding misalignment compensation, i.e.: sub-function (F_2), the most recommended approach [43] lies in the adoption of *Wood's eutectic metal*: one specimen end is inserted into a specifically designed heated pot, containing the *Wood's metal* in the liquid state, in a perfectly aligned position. Afterwards, the metal is returned to its solid state and consequently the specimen is frozen in the aligned position. However, this fixture technique has some drawbacks: only misalignments due to a transverse offset between clamps can be compensated, whereas an angular misalignment cannot be cancelled. Moreover, the Wood's metal contains lead and cadmium and has toxic properties: it can be harmful if poured on skin in its molten state. Finally, the application of Wood's metal is strongly limited by its low shear strength, about 24 MPa, according to [54], which also implies significant safety problems, when high loads are applied.

Other possible solutions include a spherical coupling [55], double cardanic joints [56, 57], the use of two spherical roller thrust bearings posed at mirrored locations and the addition of members with a high bending compliance [58].

A further option, with details below, consists in the mounting of a thrust ball bearing beneath the sample head.

3. The third sub-function F_3 may be achieved by calibrated disks or wedges that fix the loading train in the axial direction, or by lowering a cylindrical member towards the specimen upper end, applying a pre-load to the head.

8.10.2 Adopted Solution

The proposed solution is sketched in Fig. 8.24, with a detail in Fig. 8.25: its main issues are summarized below.

First of all, the specimen lower end is completely fixed by a threaded coupling. Pre-compression of the horizontal specimen surface towards a punch and a cylindrical element of the machine clamp is performed here by means of a wrench. The specimen is then constrained at its upper side: a bushing is connected to the threaded head. The bushing is placed on the inverted ring of a thrust bearing with sphered housing washers, to perform compensation of both transverse offset and rotations between clamps.

The upper head pre-compression is finally achieved, by lowering a screw, having a machined head, and by then fixing it in its final position by a locknut. The described device is able to work with or without lubrication: it is however advisable to place a thin lubricant layer at the interface between the hemispherical surface and the sphered housing. The upper member must be designed, so that a sufficient space is left available for the tightening procedure by a wrench. Reference [59] provides some

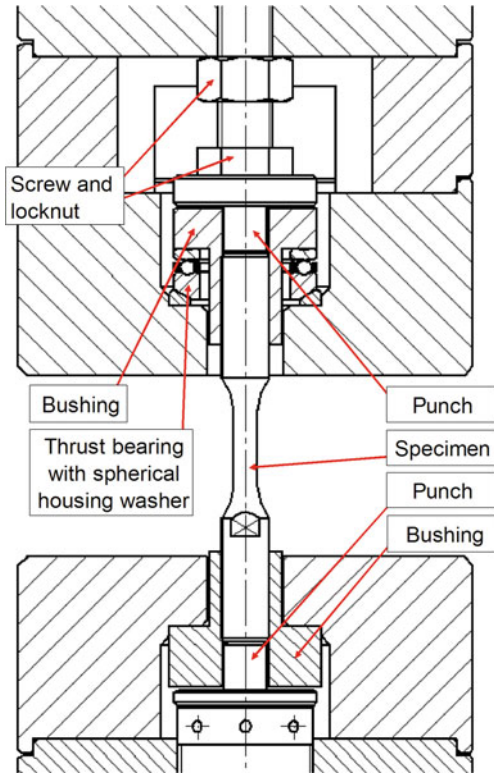


Fig. 8.24 The selected solution [53], realized by thrust bearing with sphered housing washers

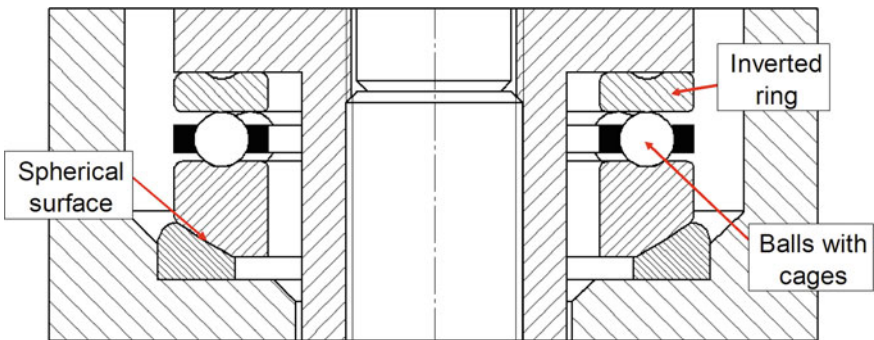


Fig. 8.25 A detail of the thrust bearing with an inverted ring

indications regarding this issue. This solution was successfully validated by experimental tests that made use of a specimen instrumented by electrical strain gages. Full details are contained in [53], while reference values for acceptable misalignments can be found in [43, 44, 60].

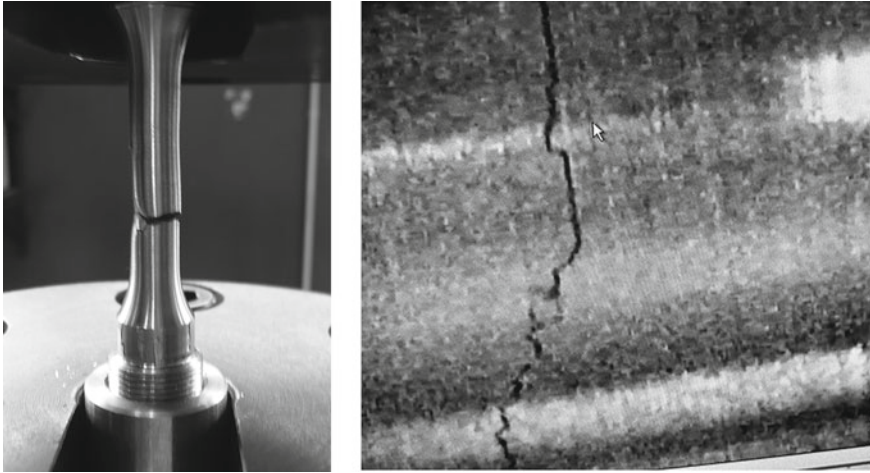


Fig. 8.26 Total fracture and first crack in a specimen axially loaded in low cycle fatigue

The specimen at the end of the test shows enucleation of cracks perpendicular to the load, circumferentially distributed, Fig. 8.26.

Many experimental results along with information regarding the experimental settings and the different options for data processing, are contained in [13, 61–66] with a recent contribution by *Atzori*.

References

1. Curioni S, Freddi A (1989) Energy based torsional fatigue analysis. In: Kussmaul (ed) 3rd international conference on multiaxialfatigue, vol 1. MPA, Universitaet of Stuttgart (MPA)
2. Haibach E (2003) FKM-guideline: analytical strength assessment of components in mechanical engineering, 5th edn. VDMA, Frankfurt
3. Halford GR, Morrow J (1962) Low cycle fatigue in torsion. In: ASTM proceedings, ASME, vol 62, pp 695–707
4. Miller J (1971) Low cycle fatigue under biaxial strain controlled conditions. *J Mater ASTM* 7(3):307–314
5. Socie DF (1977) Fatigue-life prediction using local stress-strain concepts. *Exp Mech* 17(2):50–56
6. Curioni S, Freddi A (1990) Life prediction analysis with local strain approach. *ÖIAZ Zeitschrift des Österreichische Ingenieuren und Architekten* 1(7/8)
7. Bergmann J, Seeger T (1977) Über neuere verfahren der Anrisslebensdauervorhersage für schwingbelastete Bauteile auf der Grundlage örtlicher Beanspruchungen. *Z Werkstofftech* 8-89-100 1, Verlag Chemie GmbH D, 6940 Weinheim
8. Olmi G, Freddi A (2010) Fatica oligociclica su cappe e rotor di turboalternatori. *Quaderni del DIN*, vol 1, 1st edn. Società Editrice Esculapio, Via U. Terracini 30 40131 Bologna (Italy), pp 117–142. <http://amsacta.unibo.it/2771/>
9. Wetzel RM (1977) Fatigue under complex loading. SAE Vol. AE-6, Warrendale, PA, USA
10. Imam MA, Chu HP, Rath BB (2002) Fatigue properties of titanium alloy Ti-6Al-2Cb-1Ta-0.8Mo. *Mater Sci Eng A* 323(1–2):457–461

11. Miwa Y, Jitsukawa S, Hishinuma A (1998) Development of a miniaturized hour-glass shaped fatigue specimen. *J Nucl Mater* 258–263:457–461
12. Sandhya R, Veeramani A, Bhanu Sankara Rao K, Mannan SL (1994) On specimen geometry effects in strain-controlled lowcycle fatigue. *Int J Fatigue* 16(3):202–208
13. Hirose T, Tanigawa H, Ando M, Kohyama A, Katoh Y, Narui M (2002) Radiation effects on low cycle fatigue properties of reduced activation ferritic/martensitic steels. *J Nucl Mater* 307–311
14. Kanchanomai C, Miyashita Y, Mutoh Y (2002) Low cycle fatigue behavior and mechanisms of a eutectic Sn-Pb solder 63Sn-37Pb. *Int J Fatigue* 24:671–683
15. Kanchanomai C, Miyashita Y, Mutoh Y (2002) Strain-rate effects on low cycle fatigue mechanism of eutectic Sn-Pb solder. *Int J Fatigue* 24:987–993
16. Kanchanomai C, Yamamoto S, Miyashita Y, Mutoh Y, McEvily AJ (2002) Low cycle fatigue test for solders using non-contact digital image measurement system. *Int J Fatigue* 24:57–67
17. Kanchanomai C, Miyashita Y, Mutoh Y, Mannan SL (2003) Influence of frequency on low cycle fatigue behavior of Pb-free solder 96.5Sn-3.5Ag. *Mater Sci Eng A* 345:90–98
18. Olmi G (2012) A novel method for strain controlled tests. *Exp Mech* 52(4):379–393
19. Garud YS (1981) A new approach to the evaluation of fatigue under multiaxial loadings. *Trans ASME, J Eng Mater Technol* 103:118–125
20. Stephens RI, Fatemi A, Stephens RR, Fuchs HO (2002) *Metal fatigue in engineering*, 2nd edn. Wiley, New York
21. Smith KN, Watson P, Topper TH (1970) A stress-strain function for the fatigue of metals. *J Mater (JMLSA)* 5(4):767–778
22. Matsuishi M, Endo T (1968) Fatigue of metals subjected to varying stress. Presentation report 1, Japan Society of of mechanical engineers, Fukuoka J
23. Palmgren AG (1924) Die Lebensdauer von Kugellagern (life length of roller bearings). *Zeitschrift des Vereines Deutscher Ingenieure (VDI Zeitschrift)* 68(14):339–341
24. Hoffmann K (1989) An introduction to measurements using strain gauges, vol 1, 1st edn. HBM GmbH, Darmstadt
25. Olmi G (2014) An experimental investigation on a crack propagating from a geartrain housing in an asphalt milling machine. *Eng Failure Anal* 38:38–48
26. McDiarmid DL (1991) A general criterion for high cycle multiaxial fatigue failure. *Fatigue Fract Eng Mater Struct* 14(4):429–453
27. Socie DF, Marquis GB (1999) *Multiaxial Fatigue*. Society of Automotive Engineers Inc. (SAE), Troy, Michigan. ISBN: 978-0768004533
28. Fleeter S, Zhou C, Houstis EN, Rice JR (1999) Fatigue life prediction of turbomachine blading
29. Rao J (2000) Turbine blade life estimation. Alpha Science Int'l Ltd, Oxford
30. Hutchinson JW (1968) Singular behaviour at the end of a tensile crack in a hardening material. *J Mech Phys Solids* 16:13–31
31. Mendelson A (1968) *Plasticity: theory and applications*, vol 1. Macmillan Company, New York
32. Nadai A (1950) *Theory of flow and fracture of solids*. McGraw Hill, New York
33. Yokobori Y (1968) An interdisciplinary approach to fracture and strength of solids, vol 1. Nordhoff W, Groningen, The Netherlands
34. Kikukawa M, Ohij K, Kotani S, Yokoi T (1972) A comparison of the axial and reverse—torsional strain cycling low-cycle fatigue strength of several structural materials. *Bull J ASME* 15(86):898–899
35. Gonyea DC (1973) Method for low-cycle fatigue design including biaxial stress and notch effects. *ASTM STP* 520:678–687
36. Stowell EZ (1956) Poisson's ratio and volume changes for plastically orthotropic material. *NACA TN* (3736)
37. Tanaka K, Matsuoka S, Kouzu F, Nagata K (1983) Fatigue strength of a rotor steel subjected to torsional loading simulating that occurring due to circuit breaker reclosing in an electric power plant. *Fatigue Eng Mater Struct* 6(2):103–120
38. Beste A, Friedl KH, Mayer KH, Wiemann W (1982) Ermittlung der betriebsfestigkeit von turbogeneratorenwellen infolge elektrischer storfalle. *VGB Kraftwerkstechnike* 62

39. Dowling NE (1978) Stress strain analysis of cyclic plastic bending and torsion. *J Eng Mater Technol Trans ASME* 100:294–302
40. Curioni S, Freddi A, Caligiana G, Gallano G (1983) Fatigue life prediction for shafts under torsional loading. In: Le May I and Monteiro SN (eds) *Proceeding international congress fracture prevention in energy and transport systems, COBEM Brazilian congress on mechanical engineering, Rio de Janeiro, vol 1*, pp 1–3. EMAS
41. Curioni S, Freddi A, Caligiana G (1990) Experimental techniques in torsional fatigue testing. In: SEM (ed) *SEM spring conference on experimental mechanics experimental mechanics, society for experimental mechanics, Inc., vol 1*. Bethel, Albuquerque NM (USA), see also <http://www.sem.org/>
42. Freddi A, Curioni S, Caligiana G (1989) La vita a fatica torsionale negli alberi dei turbo gruppi. La collaborazione ENEA industria nel settore termo-elettromeccanico, other collaborators: Dal Re V, Cammarota GP, Cesari F, Maggiore A, Medri GL, Nicoletto G, Persiani F, Veschi D
43. ASTM E606/E606M–12 (2012) Standard test method for strain-controlled fatigue testing. ASTM International, West Conshohocken, PA, USA
44. ISO 12106:2003 (2003) Metallic materials—fatigue testing—axial-strain-controlled method. International Organization for Standardization (ISO), Geneva, Switzerland
45. ISO 12107:2012 (2012) Metallic materials—fatigue testing—statistical planning and analysis of data. International Organization for Standardization (ISO), Geneva, Switzerland
46. EN ISO 6892-1:2009 (2009) Metallic materials—tensile testing. Part 1: Method of test at room temperature (ISO 6892-1:2009)
47. Hähner P, Rinaldi C, Bicego V et al (2008) Research and development into a European code-of-practice for strain-controlled thermo-mechanical fatigue testing. *Int J Fatigue* 30:372–381
48. Roessle ML, Fatemi A (2000) Strain-controlled fatigue properties of steels and some simple approximations. *Int J Fatigue* 22:495–511
49. Kandil FA, Dyson BF (1993) The influence of load misalignment during uniaxial low-cycle fatigue testing-I: modelling. *Fatigue Fract Eng, Mater Struct* 16(5):509–527
50. Kandil FA, Dyson BF (1993) The influence of load misalignment during uniaxial low-cycle fatigue testing-II: applications. *Fatigue Fract Eng, Mater Struct* 16(5):529–537
51. Thomas GB, Varma RK (1992) Evaluation of low cycle fatigue test data in the BCR/VAMAS intercomparison programme. In: Loveday M, Gibbons T (eds) *Harmonisation of testing practice for high temperature materials*. Elsevier Science Publisher, Oxford, pp 155–185
52. Pahl G, Beitz W (1997) *Engineering design, a systematic approach*. Springer, Berlin
53. Olmi G (2011) A new loading-constraining device for mechanical testing with misalignment auto-compensation. *Exp Tech* 35(6):61–70
54. Blomquist RS (1995) The design of a benign fail-safe mechanism using a low-melting-point metal alloy coupler. Appears in collection JPL TRS 1992, National Aeronautics and Space Administration (NASA), Houston, TX, United States. <http://hdl.handle.net/2014/33266>
55. Christ BW (1973) Effects of misalignment on the pre-macroyield region of the uniaxial stress-strain curve. *Metall Trans* 4(8):1961–1965
56. O’Kane RJ (1971) Discussion in elevated temperature testing problem areas. *Am Soc Test Mater, STP* 488:61–64
57. Penny RK, Ellison EG, Webster GA (1966) Specimen alignment and strain measurement in axial creep tests. *Mater Res Stand* 6(2):76–84
58. Birkbeck G, Petch NJ, Rae DM (1972) Eccentricity of loading in tensile testing. *J Iron Steel Inst* 210(9):675–679
59. UNI 6761–1992 (1992) Assembly tools. Minimum free space for tightening
60. Schmieder AK, Henry AT (1971) Axiality measurements on fifty creep machines. *Am Soc Test Mater, STP* 488:43–60
61. Atzori B, Meneghetti G, Ricotta M (2014) Unified material parameters based on full compatibility for low-cycle fatigue characterisation of as-cast and austempered ductile irons. *Int J Fatigue* 68:111–122
62. Olmi G (2012) An efficient method for the determination of the probability of failure on the basis of LCF data: application to turbogenerator design. *SDHM Struct Durab Health Monit* 8(1):61–89

63. Olmi G (2012) Low cycle fatigue experiments on turbogenerator steels and a new method for defining confidence bands. *J Test Eval (JTE)* 40(4):539–552
64. Olmi G, Freddi A (2014) Reliability assessment of a turbogenerator coil retaining ring based on low cycle fatigue data. *Arch Mech Eng* 61(1):5–34
65. Pang JHL, Xiong BS, Low TH (2004) Low cycle fatigue study of Lead free 99.3Sn-0.7Cu solder alloy. *Int J Fatigue* 26:865–872
66. Petersen C, Schmitt R, Gamier D (1996) Thermal and isothermal low cycle fatigue of manet i and ii. *J Nucl Mater* 233–237:285–288

Chapter 9

Elementary Models for Stress Singularities

Abstract Fracture Mechanics began in the first decade of the 20th century as the science oriented to enlighten the state of stress and strain near a crack in a stress field, in order to explain structural collapses that were not clarified by classical mechanics. Some catastrophic cases of rupture stimulated the development of this discipline, which is strongly connected to fatigue theories. After a pioneering work, the discipline flourished due to the collective work of a great number of scholars and research centers and today it accompanies every interpretation of fracture and fatigue phenomena. Experimental stress analysis offers appropriate tools for material characterization (i.e. its capability to tolerate the presence of a crack) and for establishing design methodologies to prevent fracture and to assess the safety of structures. In this chapter the use of strain gages, photoelasticity and holographic interferometry is applied to case studies of Stress Intensity Factors determinations for several kinds of cracks in structures encountered in laboratory and industrial applications.

9.1 Behavior Classification

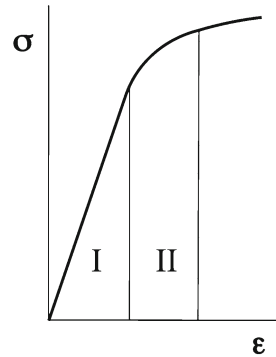
A large thin plate with a through-thickness crack is loaded by a uniform stress at a certain distance, perpendicularly to the crack plane. If the load increases monotonically, the relationship between load and crack mouth opening displacement is shown in Fig. 9.1. The diagram has the typical trend of the static σ - ε curve: for low loads the diagram has a linear behavior, even if a limited plasticity occurs at the crack tip, while additional loading will cause response to deviate from a linear behavior when a limited but larger plastic area takes place at the crack tip.

- Region I is the field of *Linear Elastic Fracture Mechanics* LEFM, where plastic deformation is confined to the crack tip region and is small enough (Small Scale Yielding) to be determined through yield criteria of Von Mises and Tresca in an elastic environment.

The stress calculation is possible using the *Irwin-Westergaard* model with the concept of *Stress Intensity Factor*.

- In region II the thin plate is subject to plastic deformation before the crack propagates. In this case the plastic crack tip zone is too large to be dealt with an elastic

Fig. 9.1 Load versus crack mouth opening displacement in a thin large rectangular plate of ductile material, loaded perpendicularly to the crack plane



theory. The material behavior is presumed to be properly nonlinear but elastic, i.e. the unloading curve of the so called elastic-plastic behavior follows the original loading curve without hysteresis loop. This assumption is a simplification for metals and an exact hypothesis for other nonlinear elastic material (e.g. polymers and composites). For this *Elastic-Plastic Fracture Mechanics* (EPFM) special models are derived that will not be considered in the present elementary handling: e.g., the *CTOD* parameter suggested by Wells and the *J Integral* proposed by Rice.¹

- This presentation does not take into account a next region of the diagram for higher loads, when the plastic zone is very large and the plate becomes totally plasticized. For a further discussion see [1–10].

9.1.1 Elastic Behavior

The analysis of a crack is obviously connected to the propagation risk that can vary between two extremes: on the one hand an instantaneous collapse at a much lower static stress than the classic admissible one, with a net parts separation (*brittle fracture*), on the other hand stable propagation for cyclic loads.

The fracture behavior is influenced by several parameters, as metallurgical variables, such as steelmaking practice and inclusions, heat treatments and microstructure, temperature, stress type (static, dynamic), residual stress and environment effects.

In a simplified phenomenological theory, all these variables are summarized in a material parameter that represents the capacity to offer resistance to a crack propagation, called *Fracture Toughness*. This parameter is compared with a structural parameter, the *Stress Intensity Factor* that is a function of the working stress, as in classical strength theory, but also of the crack shape and size.

¹ Shih showed that between *J* and CTOD a relationship exists for a given material. Thus, these two parameters are both valid in characterizing crack tip toughness for elastic-plastic materials.

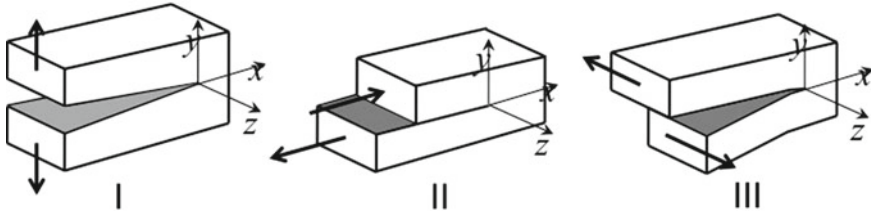


Fig. 9.2 Three modes of crack loading, opening, shearing and tearing

9.1.2 Crack Loading Modes

Irwin was one of the first to study the behavior of cracks, with reference to three different loading modes.

If a through-thickness crack is placed in a loaded body, in the most general condition it is subjected to three overlapped basic types of loading, shown in Fig. 9.2.

- The opening mode I, due to loads that produce displacements of the crack surfaces perpendicular to the plane of the crack. The dominant stress is σ_y .
- The shearing mode II, due to in-plane shear loads which cause the two crack surfaces to slide on one another. The dominant stress is τ_{yx} .
- The tearing mode III, due to out-of-plane shear loadings. The crack surfaces slide in a direction parallel to the tip front. The dominant stress is τ_{yz} .

9.1.3 Plane Stress and Plane Strain

Plane Stress There is no stress on a certain plane assumed as principal plane (1, 2-plane) and consequently all the stress components are parallel to that plane.

Plane Strain There is no elongation and shear in/ and perpendicular to one direction assumed as the 3-direction, perpendicular to the 1, 2-plane.

These two distributions are the limit cases of a weak and a strong stress three-dimensionality (that means respectively two prevalent values (e.g. σ_1, σ_2) or three similar values of the principal stress (σ_1, σ_2 and σ_3) which happen when the cracked body is thin or when it is thick (comparing its thickness to the crack size), or, even in a thick body, on the outer surface (plane stress), and in the middle plane of the thickness (plane strain). For e.g. the case in Fig. 9.1 is supposed in a plane stress state, because the plate is thin, then:

$$\sigma_z = \tau_{xz} = \tau_{yz} = 0$$

while the case of:

$$\tau_{xz} = \tau_{yz} = 0 \quad \varepsilon_z = 0$$

then $\sigma_z = \nu(\sigma_x + \sigma_y)$ that is the condition of a plane strain state.

9.2 The Field Equations for the Region Close to the Crack Tip

For modeling the singularity in the stress field due to a crack, defined as a geometrical discontinuity with a crack tip radius equal to zero, the theoretical modeling starts from the case of an elliptical hole with a semi-diameter a much longer than the semi-diameter b .

With reference to the analytical elastic solution for an elliptic hole of this type in a large plate loaded by a uniform state of stress [11], it is possible to demonstrate, introducing an elliptical coordinates system, that the maximum stress tangent to the hole and perpendicular to the larger diameter is, Fig. 9.3:

$$\sigma_{max} = 2\sigma_n \frac{a}{b} \tag{9.1}$$

where:

a = larger semi-diameter

b = shorter semi-diameter

σ_n = stress at large distance perpendicular to the longer diameter

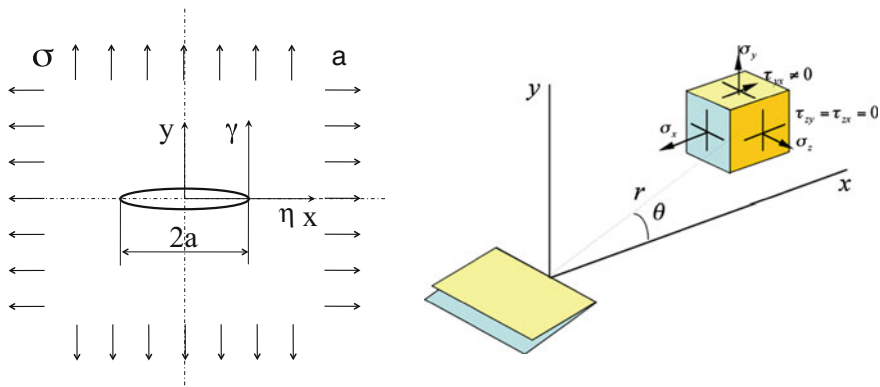


Fig. 9.3 A crack in a plate with the stress distribution at a point of the plane close to the crack tip

For a uniform stress in one direction only, perpendicular to the larger diameter, the following holds:

$$\sigma_{max} = \sigma_n \left(1 + 2\frac{a}{b} \right) \tag{9.2}$$

A crack is considered an elliptical hole with $a \gg b$, i.e. with a curvature radius at the extreme of the larger diameter tending to zero. With this assumption Eq. 9.2 tends to 9.1.

Equivalent to the previous one, another expression for maximum stress at the extreme of the larger diameter, as a function of the curvature radius ρ , is:

$$\sigma_{max} = \sigma_n \left(1 + 2\sqrt{\frac{a}{\rho}} \right) \tag{9.3}$$

where: a is the larger semi-diameter. For a radius value tending to zero, as in a real crack, Fig. 9.3 on the left, holds:

$$\sigma_{max} \approx 2\sigma_n \sqrt{\frac{a}{\rho}} \propto \sigma_n \sqrt{a} \tag{9.4}$$

for $a \gg \rho$.

If ρ tends to zero the maximum stress tends to infinity, i.e. a singularity in the stress field arises and the *Stress Concentration Factor* loses its meaning.

The present analytical approach is specific to a biaxially loaded infinite plate, Fig. 9.3. The theory starts from the general equations seen in Chap. 6, Sect. 6.3.1.

9.2.1 The Westergaard Stress Function

The earliest work dates back to Inglis, and other important developments were made by Griffith in 1920, Westergaard in 1939.

For our specific problem of an infinite, biaxially loaded plate containing a crack, *Westergaard* suggested the following way, introducing a ϕ function as [12, 13]:

$$\phi = \text{Re } \bar{\bar{Z}}(z) + y \cdot \text{Im } \bar{Z}(z) \tag{9.5}$$

where $Z(z)$ is a complex function of the complex variable $z(x + iy)$, analytic on the domain:

$$Z(z) = \text{Re } Z + i\text{Im } Z \tag{9.6}$$

and

$$z = x + iy = re^{i\theta} \quad (9.7)$$

The functions $\overline{\overline{Z}}(z)$ and $\overline{Z}(z)$ are the integrals of the first and second order.

$$\frac{d\overline{\overline{Z}}}{dz} = \overline{\overline{Z}}(z) \quad \frac{d\overline{Z}}{dz} = Z(z) \quad \frac{dZ}{dz} = Z'(z) \quad (9.8)$$

or:

$$\overline{\overline{Z}} = \int \overline{\overline{Z}}(z) dz \quad \overline{Z} = \int Z(z) dz \quad Z = \int Z' dz$$

The function is analytic since it satisfies the condition of Cauchy-Riemann. From Eq. 9.8:

$$\frac{\partial \overline{\overline{Z}}}{\partial x} = \frac{d\overline{\overline{Z}}}{dz} \cdot \frac{\partial z}{\partial x} = Z \quad \frac{\partial \overline{\overline{Z}}}{\partial y} = \frac{d\overline{\overline{Z}}}{dz} \cdot \frac{\partial z}{\partial y} = i Z \quad (9.9)$$

Thus:

$$\frac{\partial \operatorname{Re} \overline{\overline{Z}}}{\partial x} = \operatorname{Re} \frac{d\overline{\overline{Z}}}{dz} = \operatorname{Re} Z$$

$$\frac{\partial \operatorname{Re} \overline{\overline{Z}}}{\partial y} = \operatorname{Re} \frac{d\overline{\overline{Z}}}{dz} = -\operatorname{Im} Z$$

$$\frac{\partial \operatorname{Im} \overline{\overline{Z}}}{\partial x} = \operatorname{Im} \frac{d\overline{\overline{Z}}}{dz} = \operatorname{Im} Z$$

$$\frac{\partial \operatorname{Im} \overline{\overline{Z}}}{\partial y} = \operatorname{Im} \frac{d\overline{\overline{Z}}}{dz} = \operatorname{Re} Z$$

and:

$$\frac{\partial \operatorname{Re} \overline{\overline{Z}}}{\partial x} = \frac{\partial \operatorname{Im} \overline{\overline{Z}}}{\partial y} \quad \frac{\partial \operatorname{Re} \overline{\overline{Z}}}{\partial y} = -\frac{\partial \operatorname{Im} \overline{\overline{Z}}}{\partial x} \quad (9.10)$$

or:

$$\nabla^2 \operatorname{Re} Z = \nabla^2 \operatorname{Im} Z = 0 \quad (9.11)$$

This result shows that the Westergaard stress function satisfies the biharmonic equation Eq. 6.22 in Chap. 6 with the real and imaginary parts as well as with the products of these with the variables x and y . In other words the analytical function Eq. 9.5 can be used as Airy function.

It is now possible to find with some steps² the (obviously real) expressions of the stresses $\sigma_x, \sigma_y, \tau_{xy}$, differentiating the Eq.9.5 according to the Eq.9.12:

$$\begin{cases} \sigma_x = \text{Re } Z(z) - y \text{Im } Z'(z) \\ \sigma_y = \text{Re } Z(z) + y \text{Im } Z'(z) \\ \tau_{xy} = -y \text{Re } Z'(z) \end{cases} \tag{9.13}$$

Equations 9.13 yield stresses for the functions $Z(z)$, while Eq. 9.5 correctly accounts for the stress singularity at the crack tip, but the Airy function must be selected to satisfy boundary conditions of the investigated problem.

The boundary conditions for a flat plate with uniform biaxial load, equal to σ_∞ , are the following:

$$\begin{cases} \sigma_y = 0 & \text{for } -a < x < +a \text{ and } y = 0 \\ \sigma_y \rightarrow \sigma_\infty & \text{for } x \rightarrow \pm\infty \text{ and any } y \\ \sigma_y \rightarrow \infty & \text{for } x = \pm a \end{cases}$$

An example of a complex analytic function Z that satisfies these boundary conditions is the following:

$$Z(z) = \frac{\sigma}{\sqrt{1 - \frac{a^2}{z^2}}} \tag{9.14}$$

One can indeed verify that along the axis x , i.e. for $y = 0$, from the second of Eq.9.13, $\sigma_y = \text{Re } Z(z)$ holds and from Eq.9.14, for $-a < x < +a$.

$\text{Re } Z(z)$ is equal to 0. Thus $\sigma_y = 0$, as the first boundary condition requires.

For x tending to $\pm\infty$ and for any value of y , i.e. for $z \rightarrow \infty$, $Z(z)$ tends to σ_∞ and for $y = 0$ and $x = \pm a$, z^2 becomes a^2 and $Z(z)$ tends to infinity and also σ_y , in this way fulfilling the condition of a singularity.

For the present purpose it is more convenient to relocate the origin of the coordinates in the crack tip; defining with η and γ the coordinates with respect to the new reference system, the following holds:

² For e.g. for the first equation, utilizing the definitions of Eq.9.9 and the Cauchy-Riemann theorem, 9.10 (and in a similar way for the other stress components):

$$\begin{aligned} \sigma_x &= \frac{\partial^2 \phi}{\partial y^2} = \frac{\partial}{\partial y} \left[\frac{\partial}{\partial y} (\text{Re } \bar{Z}(z) + y \cdot \text{Im } \bar{Z}(z)) \right] = \frac{\partial}{\partial y} \left[\frac{\partial \text{Re } \bar{Z}(z)}{\partial z} \frac{\partial z}{\partial y} + \text{Im } \bar{Z}(z) + y \frac{\partial \text{Im } \bar{Z}(z)}{\partial y} \frac{\partial z}{\partial y} \right] \\ &= \frac{\partial}{\partial y} [i \text{Re } \bar{Z} + \text{Im } \bar{Z} + i y \text{Im } Z] = i \frac{\partial \text{Re } \bar{Z}}{\partial y} + \frac{\partial \text{Im } \bar{Z}}{\partial y} + i \text{Im } Z + y i \cdot i \text{Im } Z' \\ &= -i \text{Im } Z + \text{Re } Z - i \text{Im } Z - y \text{Im } Z' = \text{Re } Z - y \text{Im } Z'. \end{aligned} \tag{9.12}$$

$$\eta = x - a$$

$$\gamma = y$$

Then the new complex variable η is $\eta = (x - a) + iy = (x + iy)a = z - a$, from which the expression of Z is derived as function of the new variable $z = \eta + a$:

$$Z(\eta) = \frac{\sigma}{\sqrt{1 - a^2/(a + \eta)^2}} = \frac{\sigma(a + \eta)}{\sqrt{(a + \eta)^2 - a^2}} \quad (9.15)$$

Now let us consider a small area near the crack tip, where $\eta \ll a$.

With this restriction the previous equation can be approximated in this way³:

$$Z(\eta) = \frac{\sigma a}{\sqrt{2a\eta}} = \frac{\sigma\sqrt{a}}{\sqrt{2\eta}} \quad (9.16)$$

This expression is the first term of a *series expansion* of Eq. 9.15.

It is also advisable to make a change of variables, expressing the complex variable η , with origin at the crack tip, with exponential notation:

$$\eta = (x - a) + iy = r \cdot e^{i\theta}$$

where r , modulus of the complex number, is:

$$r = \sqrt{(x - a)^2 + y^2}$$

that is the distance of the point (x, y) from the new origin at the crack tip, Fig. 9.3. Equation 9.16 now becomes:

$$Z(\eta) = \frac{\sigma\sqrt{a}}{\sqrt{2\eta}} = \frac{\sigma\sqrt{a}}{\sqrt{2r} \cdot e^{i\theta/2}} = \frac{\sigma\sqrt{\pi a}}{\sqrt{2\pi r}} \cdot e^{-(1/2)i\theta} \quad (9.17)$$

The real and imaginary parts of the stress expressions, Eq. 9.13 are:

$$\begin{aligned} \operatorname{Re} Z(\eta) &= \frac{\sigma\sqrt{\pi a}}{\sqrt{2\pi r}} \cdot \cos \theta/2 \\ \operatorname{Im} (Z') &= \operatorname{Im} \frac{\sigma\sqrt{\pi a}}{\sqrt{2\pi}} \cdot \frac{d}{d\eta} \frac{1}{\sqrt{\eta}} = \frac{\sigma\sqrt{\pi a}}{\sqrt{2\pi}} \operatorname{Im} \left(\frac{-1}{2} \right) \eta^{-3/2} \\ &= -\frac{\sigma\sqrt{\pi a}}{2\sqrt{2\pi}} r^{-3/2} \sin \frac{3}{2}\theta = \frac{\sigma\sqrt{\pi a}}{2\sqrt{2\pi r}} (1/r) \cdot \sin \frac{3}{2}\theta \end{aligned} \quad (9.18)$$

Equation 9.13 becomes:

³ $\sqrt{(a + \eta)^2 - a^2} = \sqrt{\eta^2 + 2\eta a} = \sqrt{\eta(\eta + 2a)} \simeq \sqrt{2a\eta}$.

$$\sigma_x = \operatorname{Re} Z - y \cdot \operatorname{Im} Z' = \frac{\sigma\sqrt{\pi a}}{\sqrt{2\pi r}} \cdot \left[\cos \frac{\theta}{2} - y \frac{1}{2r} \sin \frac{3}{2}\theta \right]$$

since: $y \frac{1}{2r} = \frac{1}{2} \sin \theta$

$$\sigma_x = \frac{\sigma\sqrt{\pi a}}{\sqrt{2\pi r}} \cdot \left[\cos \frac{\theta}{2} - \frac{1}{2} \sin \theta \sin \frac{3}{2}\theta \right]$$

Remembering that:

$$\frac{1}{2} \sin \theta = \sin \frac{\theta}{2} \cos \frac{\theta}{2}$$

σ_x becomes:

$$\sigma_x = \frac{\sigma\sqrt{\pi a}}{\sqrt{2\pi r}} \cos \frac{\theta}{2} \left[1 - \sin \frac{\theta}{2} \sin \frac{3}{2}\theta \right] \tag{9.19}$$

In a similar way, the expressions of the other stress components are derived.

Concluding, with the approximation of the first term of a series expansion (justified by the condition of $r \ll a$), Eq. 9.15 can be approximated by a simplified expression of the stress function Z Eq. 9.19, that leads to the following equations:

$$\begin{aligned} \sigma_x &= \frac{K_I}{\sqrt{(2\pi r)}} \cos \frac{\theta}{2} \left(1 - \sin \frac{\theta}{2} \sin \frac{3}{2}\theta \right) \\ \sigma_y &= \frac{K_I}{\sqrt{(2\pi r)}} \cos \frac{\theta}{2} \left(1 + \sin \frac{\theta}{2} \sin \frac{3}{2}\theta \right) \\ \tau_{xy} &= \frac{K_I}{\sqrt{(2\pi r)}} \cos \frac{\theta}{2} \sin \frac{\theta}{2} \cos \frac{3}{2}\theta \end{aligned} \tag{9.20}$$

$$\tau_{xz} = \tau_{yz} = 0 \quad \sigma_z = 0 \quad \text{for Plane Stress}$$

$$\sigma_z = \nu (\sigma_x + \sigma_y) \quad \text{for Plane Strain.}$$

and ν is the *Poisson's* ratio.

All the stresses tend to infinity at the crack tip (for $r = 0$) and can be expressed by the product of $\sigma\sqrt{\pi a}$ and other standard terms as $1/\sqrt{2\pi r} f(\theta)$.

The term $\sigma\sqrt{\pi a}$ is a function of applied stress, i.e. of the load, and of the square root of the half length of the crack. This factor, indicated with K_I , is called *Stress Intensity Factor* (SIF), because it is responsible for the *intensification* of the stress at the crack tip and then of its possible propagation.

From the equilibrium equations, the corresponding relationships for the principal stresses in mode I can be derived:

$$\begin{cases} \sigma_1 = \frac{K_I}{\sqrt{2\pi r}} \cos \frac{\theta}{2} \left(1 + \sin \frac{\theta}{2}\right) \\ \sigma_2 = \frac{K_I}{\sqrt{2\pi r}} \cos \frac{\theta}{2} \left(1 - \sin \frac{\theta}{2}\right) \\ \sigma_3 = \frac{2\nu K_I}{\sqrt{2\pi r}} \cos \frac{\theta}{2} \end{cases} \quad (9.21)$$

The previous formulas are conceptually justified observing that σ_y , for r and a tending to zero, tends to a finite value that must be the nominal stress σ . Since:

$$\sigma_y(r, a) \propto \frac{\sigma \sqrt{a}}{\sqrt{r}}$$

we have:

$$\lim_{a \rightarrow 0, r \rightarrow 0} \sigma_y(r, a) = \frac{0}{0}.$$

9.2.2 Displacements Close to the Crack Tip

The presentation is limited to the biaxial case that will be illustrated by an experiment, remembering the relationship between stress and strain in this case:

$$\begin{aligned} \varepsilon_x &= \frac{1}{E}(\sigma_x - \nu\sigma_y) \\ \varepsilon_y &= \frac{1}{E}(\sigma_y - \nu\sigma_x) \\ \tau_{xy} &= \frac{1}{G}\tau_{xy} \end{aligned} \quad (9.22)$$

Integrating the equations like this:

$$\varepsilon_x = \frac{\partial u}{\partial x}$$

the following relationships are obtained for the Plane Stress state:

$$u = 2 \frac{K_I}{E} \sqrt{\frac{r}{2\pi}} \cos \frac{\theta}{2} \left(1 + \sin^2 \frac{\theta}{2} - \nu \cos^2 \frac{\theta}{2}\right) \quad (9.23)$$

and a similar one for the other displacement components:

$$v = 2 \frac{K_I}{E} \sqrt{\frac{r}{2\pi}} \sin \frac{\theta}{2} \left(1 + \sin^2 \frac{\theta}{2} - \nu \cos^2 \frac{\theta}{2}\right) \quad (9.24)$$

Similarly for the Plane Strain state:

$$u = \frac{K_I}{G} \sqrt{\frac{r}{2\pi}} \cos \frac{\theta}{2} \left(1 - 2\nu + \sin^2 \frac{\theta}{2} \right) \tag{9.25}$$

$$v = \frac{K_I}{G} \sqrt{\frac{r}{2\pi}} \sin \frac{\theta}{2} \left(2 - 2\nu - \cos^2 \frac{\theta}{2} \right) \tag{9.26}$$

where $G = \frac{E}{2(1 + \nu)}$.

As a particular case, referring to Eq. 9.26, for $\theta = \pi$ the following holds:

$$\nu = \frac{2K_I}{G} \sqrt{\frac{r}{2\pi}} (1 - \nu).$$

9.2.3 Stress Field for Modes II and III

In a similar way, the local stresses near the crack tip can be expressed for the stress fields of Mode II and Mode III.

For Mode II: ($K_{II} = \tau \sqrt{\pi a}$):

$$\begin{aligned} \sigma_x &= \frac{-K_{II}}{\sqrt{2\pi r}} \sin \frac{\theta}{2} \left(2 + \cos \frac{\theta}{2} \cos \frac{3}{2}\theta \right) \\ \sigma_y &= \frac{K_{II}}{\sqrt{2\pi r}} \sin \frac{\theta}{2} \left(\cos \frac{\theta}{2} \cos \frac{3}{2}\theta \right) \\ \tau_{xy} &= \frac{K_{II}}{\sqrt{2\pi r}} \cos \frac{\theta}{2} \left(1 - \sin \frac{\theta}{2} \sin \frac{3}{2}\theta \right) \end{aligned} \tag{9.27}$$

and for Mode III: ($K_{III} = \tau \sqrt{\pi a}$):

$$\begin{aligned} \tau_{xz} &= \frac{-K_{III}}{\sqrt{2\pi r}} \sin \frac{\theta}{2} \\ \tau_{yz} &= \frac{-K_{III}}{\sqrt{2\pi r}} \cos \frac{\theta}{2} \end{aligned} \tag{9.28}$$

In the general case, the relationship between the stress state in a region near the tip and the Stress Intensity Factors is the following:

$$\sigma_{ij} = \frac{1}{\sqrt{2\pi r}} \left(K_I f_{ij}^I + K_{II} f_{ij}^{II} + K_{III} f_{ij}^{III} \right) \quad (9.29)$$

where K_I , K_{II} , K_{III} are the Stress Intensity Factors for the three Modes.

9.2.4 Magnification Factor

In a small region near the crack the stress field generated by a crack depends only on the stress intensity factors that varies as a linear function of the applied stress and of the square root of the crack length. We can say e.g. that, for Mode I the stress, σ_y at the tip reaches a critical value when K_I reaches a critical value.

For a plane crack ($\theta = 0$), holds:

$$\sigma_y = \frac{K_I}{\sqrt{2\pi r}} \quad (9.30)$$

Thus the real asymptotic meaning of the stress intensity factor is the following:

$$K_I = \lim_{r \rightarrow 0} \sigma_y \sqrt{2\pi r} \quad (9.31)$$

In the most simple case of a crack of length $2a$ in a large plate loaded by a stress σ perpendicular to the crack plane, K_I holds:

$$K_I = \sigma \sqrt{\pi a} \quad (9.32)$$

Taking into account the effect of finite boundary and of different loads distributions, for Mode I the Stress intensity factor can be expressed in the form:

$$K_I = M\sigma \sqrt{\pi a} \quad (9.33)$$

where the factor M modifies the elementary solution developed for an infinite plane under bi-axial loadings for adapting it to different boundary and loading conditions.

For three-dimensional cracks, the K -Factors must be defined for each direction perpendicular to the front of the crack tip. In the particular case of a circular defect with uniform loading perpendicular to its plane, the value of stress intensity factor K_I which is constant for all radial directions, becomes:

$$K_I = \frac{2}{\pi} \sigma \sqrt{\pi a} \quad (9.34)$$

Also in this case for Mode I the value of K is always of the form:

$$K_I = M \frac{2}{\pi} \sigma \sqrt{\pi a} \tag{9.35}$$

where: M is a multiplier, called *Magnification factor* that refers any particular case to a basic solution.

In all other three-dimensional cases, the same factor for Mode I is variable at every point of the tip front and can be determined, when possible, theoretically [14–17] or more frequently, by numerical or experimental approaches.

9.2.5 More Terms of Series Expansion

Referring again to Z function (Westergaard function), a question arises about what happens when only the first term of Taylor series of the function Z , Eq. 9.16 is not sufficient to describe the function and more terms must be taken into account.

For the sake of simplicity, the analysis is limited to the determination of the stress value σ_x along the crack axis, where $\sigma_x = \text{Re } Z$ and $\eta = r$ (being $\theta = 0$), but the same thing happens at every point in the plane.

Considering the function Z in the complete form of Eq. 9.15, the Taylor series expansion, e.g. for the first three terms, is⁴:

$$\sigma_x = \text{Re } Z(r) = \frac{\sigma \sqrt{a}}{\sqrt{2r}} + \frac{3\sigma \sqrt{r}}{4\sqrt{2a}} + \frac{5\sigma r^{3/2}}{32\sqrt{2a^3/2}} + 0(r^{5/2}) \tag{9.36}$$

By introducing the expression of SIF ($K_I = \sigma \sqrt{\pi a}$) in all the series terms, the Eq. 9.36 becomes:

$$\sigma_x = \frac{K_I}{\sqrt{2\pi r}} + \frac{3}{4} \frac{K_I}{\sqrt{2\pi r}} \frac{r}{a} + \frac{5}{32} \frac{K_I}{\sqrt{2\pi r}} \left(\frac{r}{a}\right)^2 + 0(r^{5/2}) \tag{9.37}$$

⁴ In general, the Taylor series for $f(x)$ is:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \dots + \frac{f^n(x_0)}{n!}(x - x_0)^n + o((x - x_0)^n)$$

where o is such that: $\lim_{x \rightarrow x_0} \frac{o((x - x_0)^n)}{(x - x_0)^n} = 0$

Instructions text for Mathematica® is:

```
Series[s * (a + r)/Sqrt[(a + r)^2 - a^2], {r, 0, 2}]
```

Then the following holds until the third term in the series expansion:

$$\frac{\sqrt{a}s}{\sqrt{2}\sqrt{r}} + \frac{3s\sqrt{r}}{4\sqrt{2}\sqrt{a}} - \frac{5sr^{3/2}}{32(\sqrt{2}a^{3/2})} + o(r)^{5/2}.$$

or:

$$\sigma_x = \frac{K_I}{\sqrt{2\pi r}} \left[1 + \frac{3r}{4a} + \frac{5}{32} \left(\frac{r}{a} \right)^2 \right] + O(r^{5/2}) \quad (9.38)$$

It would be erroneous to use the Eq. 9.38 for determining the K value by one measurement of σ_x at one point. It is necessary to estimate the SIF value at a large number of points by Least Squares procedure. See an example in the next Sect. 9.5.1.3.

Generalization of the three terms can be extended to each stress and strain component for any value of the angle θ [3, 12].

9.3 Identification of the Stress Intensity Factor Utilizing More Terms of the Series

Stress Intensity Factors are determined for existing cracks in order to assess the integrity of structural components. Magnification Factors have a role similar to Stress Concentration Factors in the fatigue theory of beams. The elementary case assumed as reference is the crack in an infinity body. For other cases of different geometries and loading conditions, if the analysis is circumscribed to a region near the crack tip, the singularity is yet represented by the single value of the stress intensity factor even if its value is influenced by the stress (strain) values at points not very close to the tip. A more accurate description of the stress state near the crack is thus necessary for the experimental determination of the right value of the Stress Intensity Factor.

The stress Intensity factor identification can be done modeling the theoretical expression of the stress (or strain) state with series developments of the stress functions with a number of terms adequate to be fitted to the experimental data [12]. The Westergaard function is thus modified, in order to express stresses and strains with at least three terms of a series development, plus a term $\sigma = const.$ for accounting the constant value of stress in x direction [12, 18, 19].

The determination of the stresses (or strains) must be done in a not too near but not too far region from the singularity: it would be impossible to fit data at points far from the crack tip, for the great number of series coefficients required. Not only: data taken too far from the crack tip would generate an ill-posed problem of SIF identification, since the singularity of the stress field tends to reduce rapidly its influence at point not close to the tip.

The field variables become in this hypothesis:

$$\begin{cases} \sigma_x = \operatorname{Re} Z - y \operatorname{Im} Z' - y \operatorname{Im} Y' + 2\operatorname{Re} Y \\ \sigma_y = \operatorname{Re} Z + y \operatorname{Im} Z' + y \operatorname{Im} Y' \\ \tau_{xy} = -y \operatorname{Re} Z' - y \operatorname{Re} Y' - \operatorname{Im} Y \end{cases} \quad (9.39)$$

where the stress functions are given as series relationships in term of z as:

$$\begin{cases} Z(z) = \sum_{n=0}^N A_n z^{(n-1/2)} \\ Y(z) = \sum_{n=0}^M B_n z^n \end{cases} \quad (9.40)$$

Substituting Eq. 9.40 in Eq. 9.39 gives⁵:

$$\begin{aligned} \sigma_x = & A_0 r^{-1/2} \cos \frac{\theta}{2} \left(1 - \sin \frac{\theta}{2} \sin \frac{3}{2}\theta \right) + 2B_0 \\ & + A_1 r^{1/2} \cos \frac{\theta}{2} \left(1 + \sin^2 \frac{\theta}{2} \right) \\ & + A_2 r^{3/2} \left(\cos \frac{3\theta}{2} - \frac{3}{2} \sin \theta \sin \frac{\theta}{2} \right) \end{aligned} \quad (9.41)$$

$$\begin{aligned} \sigma_y = & A_0 r^{-1/2} \cos \frac{\theta}{2} \left(1 + \sin \frac{\theta}{2} \sin \frac{3}{2}\theta \right) \\ & + A_1 r^{1/2} \cos \frac{\theta}{2} \left(1 - \sin^2 \frac{\theta}{2} \right) \\ & + A_2 r^{3/2} \left(\cos \frac{3\theta}{2} + \frac{3}{2} \sin \theta \sin \frac{\theta}{2} \right) \end{aligned} \quad (9.42)$$

$$\begin{aligned} \tau_{xy} = & A_0 r^{-1/2} \cos \frac{\theta}{2} \sin \frac{\theta}{2} \cos \frac{3}{2}\theta \\ & - A_1 r^{1/2} \sin \frac{\theta}{2} \cos^2 \frac{\theta}{2} \\ & - 3A_2 r^{3/2} \sin \frac{\theta}{2} \cos^2 \frac{\theta}{2} \end{aligned} \quad (9.43)$$

where the constant A_0 the following holds:

$$A_0 = \frac{K_I}{\sqrt{2\pi}} \quad (9.44)$$

If, for small values of r , only the first term of the series is kept and the others are neglected, Eq. 9.43 is reduced to:

$$\sigma_x = \frac{K_I}{\sqrt{2\pi r}} \cos \frac{\theta}{2} \left(1 - \sin \frac{\theta}{2} \sin \frac{3}{2}\theta \right) \quad (9.45)$$

⁵ The B_0 constant is called uniform remote stress or biaxial correction.

To increment the accuracy, selecting two terms ($n = 0, 1$) instead of one, Eq. 9.16, it is possible to take into account the constant stress σ_0 ⁶:

$$2B_0 = -\sigma \quad (9.46)$$

Augmenting the distance but remaining close to the crack tip, the number of terms of series A_i can be limited to three values (A_0, A_1, A_2) and to one value of B_m (B_0).

Similar equations can be written for the Sharing Mode II [12].

9.4 Crack Tip Plasticity

The elastic solution for a singularity in the stress field, is not able to deal with the 3-D stress state at the tip of the crack, where the yield strength is overcome and a plastic deformation occurs.

The plastic zone size must be small with respect to the crack dimension and with respect to the ligament, i.e. the thickness from the crack tip to the closest boundary, in order to validate the use of an approximate elastic solution. The plastic zone size can be analytically estimated applying a plasticity criterium to the Eqs. 9.20 and 9.21, for several values of r and θ , and for the yield condition in Plane Stress and in Plain Strain.

If the specimen is thin, the Plane Stress is present along the whole thickness and the principal stresses are $\sigma_y = \sigma_x \neq 0$ and $\sigma_z = 0$, while if the body is thick the Plane Stress condition is only at the free surfaces and the condition of Plane Strain tends to be inside the thickness.

It is necessary to distinguish the plastic zone size along the crack direction ($\theta = 0$) from the plastic zone size in other directions.

9.4.1 Apparent Plastic Zone Size According to Von Mises Criterion

At the tip of the crack and in a very close region near the tip, the local stress exceeds the material yield strength and the stress singularity, foreseen by the elementary elastic theory, cannot occur. In order to preserve a general elastic analysis, the size of the plastic zone must be small with respect to the ligament dimension from the crack tip to the nearest boundary.

⁶ This correction is necessary when the stress analysis is not confined only to a region very near to the crack tip, since the Westergaard model of a crack in a large sheet was theoretically developed for a uniform state of stress, i.e. for a stress along the y but also the x axis. This value does not modify the SIF calculation in the previous elementary hypothesis, but its influence can become relevant if the K factor derives from experimental data taken from a larger region around the crack tip (as it happens by optical but especially by strain gages methods).

Let the Von Mises criterion be applied at *plane* stress and strain states for describing the plastic zone shape and size at the front of the crack tip. It will be shown why there must be a difference between the concept of apparent size dimensions and true size dimensions [12].

The apparent size derives from a simple direct plasticity condition, i.e. the condition of material yielding.

The true size derives from the further condition of a global re-distribution of stresses as a consequence of the apparent plastic zone size.

9.4.1.1 Plane Stress State

Substituting, in the following condition of plasticity:

$$\sigma_{eq} = \frac{1}{\sqrt{2}} \sqrt{(\sigma_1 - \sigma_2)^2 + (\sigma_1 - \sigma_3)^2 + (\sigma_2 - \sigma_3)^2} = \sigma_Y \tag{9.47}$$

the expressions of σ_1 and σ_2 from Eq. 9.21 (in this case $\sigma_3 = \sigma_z = 0$), a condition is derived for the apparent plastic radius of the plastic zone r_{pl} :

$$r_{pl} = \frac{1}{2\pi} \left(\frac{K_I}{\sigma_Y} \right)^2 \cos^2 \frac{\theta}{2} \left(3 \sin^2 \frac{\theta}{2} + 1 \right) \tag{9.48}$$

For $\theta = 0$ i.e. along the direction of the crack, the (apparent) radius of the plastic zone is:

$$r_{pl} = \frac{1}{2\pi} \left(\frac{K_I}{\sigma_Y} \right)^2 \tag{9.49}$$

The shape of the plastic zone is described by the Eq. 9.48, and shown in Fig. 9.4 as the largest area.

9.4.1.2 Plane Strain State

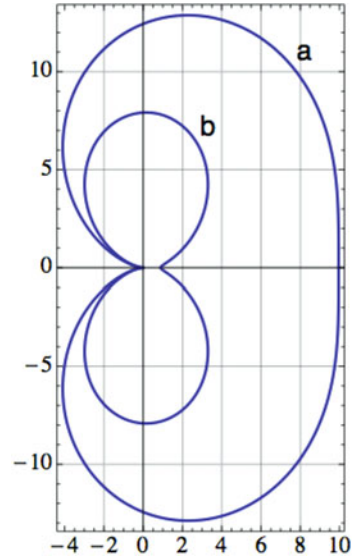
From the same Eq. 9.47 with the three Eq. 9.21, the following relationship is derived for the apparent plastic radius:

$$r_{pl} = \left(\frac{K_I}{\sigma_Y} \right)^2 \cos^2 \frac{\theta}{2} \left(3 \sin^2 \frac{\theta}{2} + (1 - 2\nu)^2 \right) \tag{9.50}$$

The shape of the plastic zone is described by the Eq. 9.50, shown in Fig. 9.4 (smallest area).

For $\theta = 0$ the radius of the plastic zone on the crack axis is:

Fig. 9.4 Example of polar plots of plastic radius in Plane Stress (*larger curve*) and in Plane Strain (*smaller curve*). Mat.: steel with $K_I = 100 \text{ MPa}\sqrt{\text{m}}$, $\sigma_Y = 400 \text{ MPa}$, $\nu = 0.35$, r_{pl} and axes in (mm)



$$r_{pl} = \frac{(1 - 2\nu)^2}{2\pi} \left(\frac{K_I}{\sigma_Y} \right)^2 \quad (9.51)$$

The size depends on the Poisson coefficient value. For $\nu = 0.5$ this dimension is zero, that means on the line of crack the material behaves in a brittle way.

9.4.2 Apparent Plastic Zone Size According to the Tresca Criterion

By means of the Tresca criterion, along the crack direction only, we find the same values of the previous model: for the plane stress condition ($\sigma_3 = 0$):

$$\sigma_x = \sigma_y = \frac{K_I}{\sqrt{2\pi r}}$$

$$\sigma_1 - \sigma_2 = \sigma_y - \sigma_x = 0$$

$$\sigma_1 - \sigma_3 = \sigma_y = \sigma_x$$

$$\sigma_2 - \sigma_3 = \sigma_y = \sigma_x$$

then:

$$\sigma_{eq} = \sigma_y = \sigma_x = \frac{K_I}{\sqrt{2\pi r}} \leq \sigma_Y$$

$$r_{pl} = \frac{1}{2\pi} \left(\frac{K_I}{\sigma_Y} \right)^2 \tag{9.52}$$

while in the case of plane strain:

$$\sigma_{eq} = \sigma_y - \sigma_z = \sigma_y - \nu(\sigma_y + \sigma_x) = \sigma_y(1 - 2\nu) \leq \sigma_Y$$

$$r_{pl} = \frac{(1 - 2\nu)^2}{2\pi} \left(\frac{K_I}{\sigma_Y} \right)^2 \tag{9.53}$$

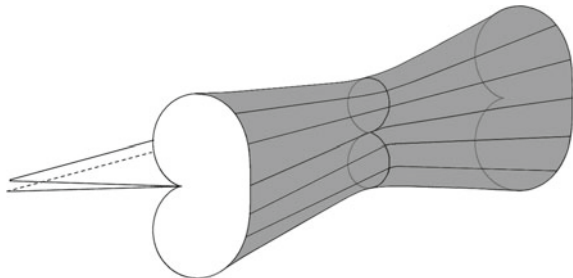
An analogy exists between the shape of the border of the plastic zone at Plane Strain state, when $\nu = 0.5$, and the shape of isochromatics around the crack tip, because both curves are loci of the difference between the two principal stresses in the plane (1–2), see Fig. 9.12 in Sect. 9.5.

As a consequence, in the cracked body with considerable thickness, the shape of the apparent plastic zone follows a form that is a larger curve at the surface, according to Fig. 9.4a, and a smaller curve in the internal part, Fig. 9.4b, with a gradual passage from one to the other shape along the specimen thickness. This typical shape shown in Fig. 9.5 is called at *dog-bone*.

This criterion indicates the direction of maximum shear stress and clarifies the reason for the thickness contraction, Fig. 9.6 in the case of plain stress: reasoning on the Mohr’s circles, the shear planes are the planes at 45° of the solid angle between the plane (1–2) and the plane (1–3).

Vice-versa, in the case of plane strain the shear planes σ are in the directions shown in Fig. 9.7. In the case of Plane Strain the plastic zone is smaller and offers a lower barrier to the crack propagation, until the value of K_I is less than a critical value K_{Ic} . When this value is reached, the fracture takes place in a brittle mode.

Fig. 9.5 *Dog-bone* plastic zone at the crack front in a thick body



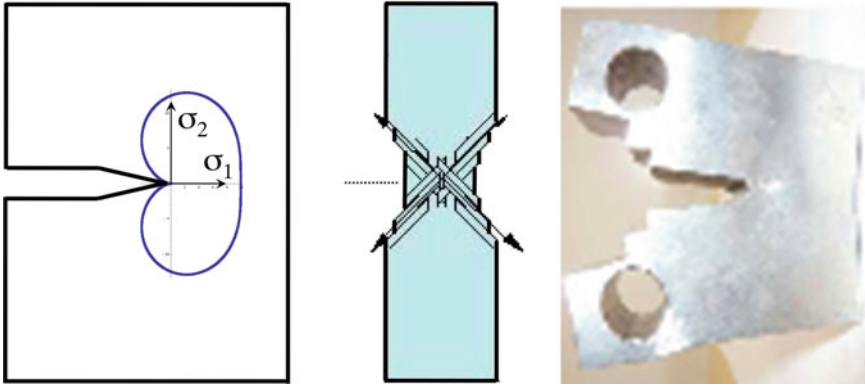


Fig. 9.6 Thickness contraction in the case of Plane Stress state in the plastic zone near the tip in a compact cracked specimen

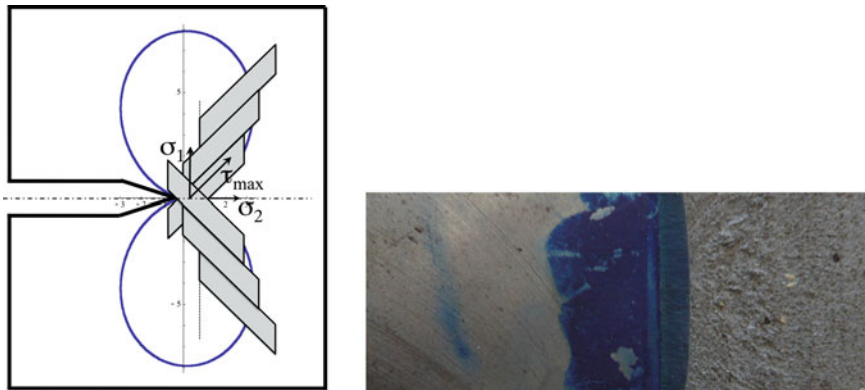


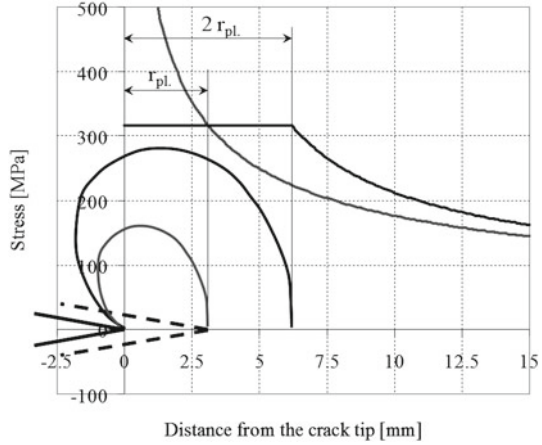
Fig. 9.7 Plane Strain state and brittle fracture of the compact specimen. The dark area was controlled with a penetrant liquid during the pre-cracking phase

9.4.3 Apparent and True Plastic Zone Size

Especially in the case of Plane Stress, once the equivalent stress has reached yield strength, a redistribution of stresses on the front of the crack is produced, in a way similar to that seen in a notch.

The equilibrium relationships between stresses and loads require that the actual radius r , and then the stress curve as function of r , is shifted by a quantity equal to r_{pl} , Fig. 9.8. With this assumption, if the plastic zone is small, the elastic solution can still be employed, taking care to hypnotize an equivalent crack of length equal to $a + r_{pl}$, where r_{pl} is the size of the plastic area along the crack axis x . In this hypothesis, it therefore has a value of K_{Ip} larger than K_I such that:

Fig. 9.8 New stress distribution (diagram on the right) at the crack tip due to plasticization, in the case of plane stress



$$K_{Ip} = M_p \cdot K_I \tag{9.54}$$

where:

$$M_p = K_{Ip}/K_I = \sqrt{1 + \frac{r_{pl}}{a}}$$

Remembering that $K_I = M\sigma\sqrt{\pi a}$ we have:

$$M_p = \sqrt{1 + \alpha \left(\frac{M \cdot \sigma}{\sigma_{pl}}\right)^2} \quad \alpha = \frac{(1 - 2\nu)^2}{2\pi} \quad \text{or} : \quad \frac{1}{2\pi} \tag{9.55}$$

From this observation it can be concluded that the true plastic zone radius is not the value of the apparent plastic zone radius r_p but its double, $2r_p$. All the observations on the shape of the plastic zone are right, but they must be considered valid for the double of the apparent plastic radius.

9.5 Case Studies of the Stress Intensity Factor Identification for Internal Cracks

The parameter $K (=M\sigma\sqrt{\pi a})$ is a combination of load and geometry data. Through a fatigue propagation of an initial crack and experimental measurements at selected propagation instants, the whole phenomenon of propagation and of stress distribution around the crack can be clarified together with the limits of the theoretical models.

9.5.1 Identification of SIF by Photoelasticity

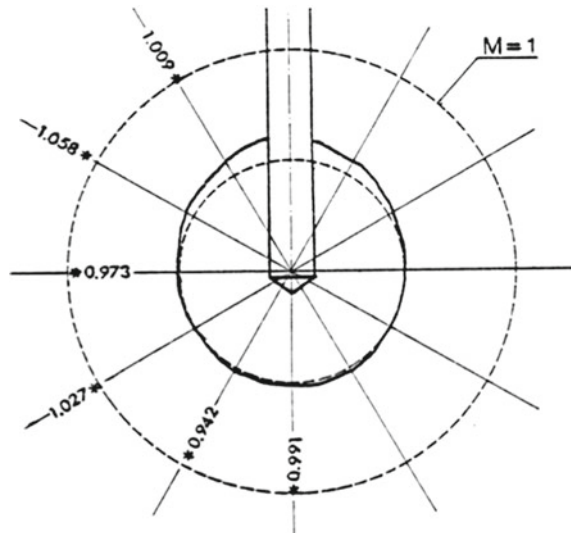
Physical models of transparent plastic (of *Araldite* epoxy resin) were utilized for simulating the crack propagation:

1. For visualizing the crack during the propagation phase.
2. For freezing the stresses at a certain instant of time and thus determining the Fracture Mechanics Parameters (as the Stress Intensity Factor SIF), in special configurations.

For generating a circular crack in a block of epoxy resin (*Araldite*), a simple procedure proved quite adequate:

- Very accurate cracks of circular form are obtained by an impact of a sharp pin inserted in a small hole, Fig. 9.9.
- The crack is *pre cracked* (propagated) by cyclic internal pressurization up to a desired front diameter size.
- Once the desired size has been reached, the crack is statically pressurized and the stresses are *frozen*, following the protocol described in Chap. 3. The stress state differs from that produced by an external uniform load by a constant value, Fig. 9.10, but the *Stress Intensity Factor* SIF is not influenced, to a first approximation, by this alteration of the load distribution and is the same in both cases.
- The *frozen* block is cut in slices along several radial directions, Fig. 9.11. The fringe pattern is shown in Fig. 9.12a.

Fig. 9.9 Circular crack obtained by an impact load on a sharp pin inserted into a small hole



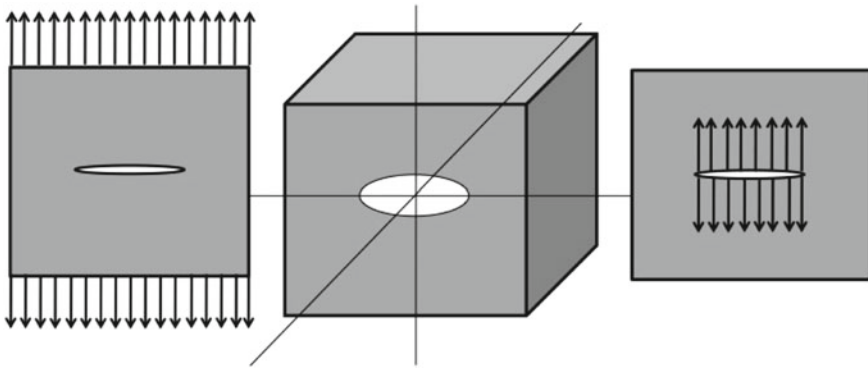


Fig. 9.10 Equivalence of SIF (Stress Intensity Factor) for internal and external loads acting on a cracked massive body

9.5.1.1 SIF Identification by the Simplest Data Extrapolation

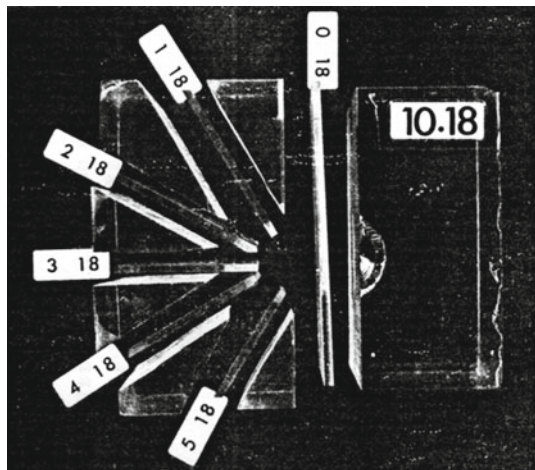
The Fracture Mechanics Equations for Mode I, give the following expression for the difference of the principal stresses in the slice plane [20]:

$$\sigma_1 - \sigma_2 = 2A \sqrt{\sin^2 \frac{\theta}{2} \sin^2 \frac{3}{2}\theta + \sin^2 \frac{\theta}{2} \cos^2 \frac{3}{2}\theta} = 2A \sin \frac{\theta}{2} \tag{9.56}$$

where:

$$A = \frac{K_I}{\sqrt{2\pi r}} \cos \frac{\theta}{2}$$

Fig. 9.11 Slicing of the block epoxy frozen model



Equation 9.56 becomes:

$$\sigma_1 - \sigma_2 = 2 \frac{K_I}{\sqrt{2\pi r}} \sin \frac{\theta}{2} \cos \frac{\theta}{2} = \frac{K_I}{\sqrt{2\pi r}} \sin \theta$$

and for $\theta = \frac{\pi}{2}$ $\sigma_1 - \sigma_2 = \frac{K_I}{\sqrt{2\pi r}} = f_\sigma \frac{N}{d}$

From this equation the following is derived:

$$\frac{K_I}{\sqrt{2\pi}} \frac{d}{f_\sigma} = N \sqrt{r}$$

or, if $C = \frac{1}{\sqrt{2\pi}} \frac{d}{f_\sigma}$:

$$C \cdot K_I = N \cdot \sqrt{r} \tag{9.57}$$

Given a fringe pattern, the elementary and simplest procedure for the determination of the Stress Intensity Factor of Mode I is based on a linear regression of data along the vertical direction, selected since this line is the richest in information, cutting numerous interferometric fringes in a brief segment, Fig. 9.12:

- After recording the fringe orders N and the relative distances on the vertical axis r from the crack tip, Fig. 9.12a, a graphics of the points of the product $N \cdot \sqrt{r}$ on the ordinate versus \sqrt{r} on the abscissa, is shown, Fig. 9.12b.
- The term $N \cdot \sqrt{r}$, for Eq. 9.57, is equal to the K_I factor, multiplied by a constant (that depends on the model thickness and on the photoelastic constant).
- The points are linearly extrapolated up to the intercept on the ordinate axis.
- The intercept is proportional to the limit of K for \sqrt{r} tending to zero, i.e. for definition, the value of K_I .

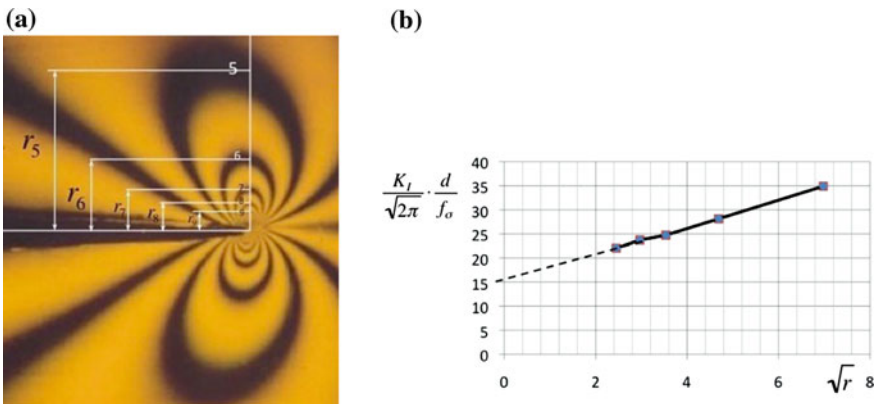


Fig. 9.12 Photoelastic determination of K_I for a circular crack in a uniform stress field by a simplified extrapolation in a single direction. **a** The interferometric image, **b** the linear extrapolation

In the case study, the straight line in Fig. 9.12b is determined for the following values of the constants:

- $E = 36 \text{ MPa}$
- $\nu = 0.45$
- $\lambda = 577 \cdot 10^{-9} \text{ m}$
- $p = 0.2 \text{ MPa}$

9.5.1.2 Errors Due to Fitting Data to a Straight Line

It must be observed that linear extrapolation of a curve for estimating the intercept is an operation subjected to numerical errors that could be unacceptable and lead to false results [21]. The following numerical results show the discrepancy between theoretical and experimental solutions.

Exercise 9.1 (Comparison between theoretical and experimentally extrapolated results) For the material data of this example, if we compare the estimated K_I value with the corresponding theoretical value, we find a significant discrepancy:

$$K_I = \frac{f\sigma}{d} \sqrt{2\pi} \cdot I = 0.688 \text{ MPa}\sqrt{\text{mm}}$$

Since $p = 0.188 \text{ MPa}$, $a = 15.11 \text{ mm}$, $d = 3 \text{ mm}$, while the theoretical value is:

$$2p\sqrt{\frac{a}{\pi}} = 0.825 \text{ MPa}\sqrt{\text{mm}}$$

The difference is about 17 %, too large and not justified by the relative accuracy of the frozen stress three dimensional photoelastic technique. A different approach should be developed for finding, with the same technique, much better results.

The error on intercept due to the extrapolation can be estimated [21]. Given the straight line of equation (Fig. 9.13):

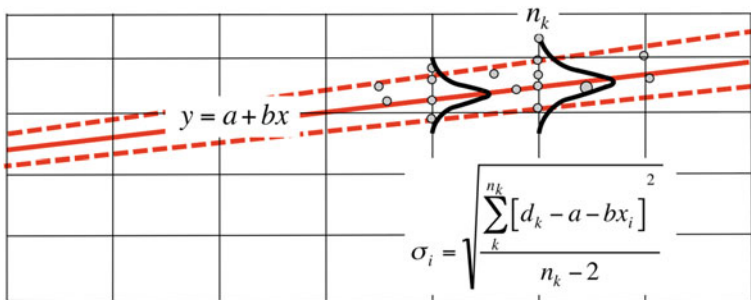


Fig. 9.13 Error on the intercept of a regression straight line in the extrapolation of experimental data

$$y = a + bx \quad (9.58)$$

the standard deviation of the intercept a and of the slope b are, respectively:

$$\begin{cases} \sigma_a = \sigma \sqrt{\frac{\sum_i \frac{x_i^2}{\sigma_i^2} / \Delta}{n_k - 2}} \\ \sigma_b = \sigma \sqrt{\frac{1}{\sum_i \frac{1}{\sigma_i^2} / \Delta}} \end{cases} \quad (9.59)$$

where:

$$\sigma_i = \sqrt{\frac{\sum_{k=1}^{n_k} [d_k - a - bx_i]^2}{n_k - 2}} \quad (9.60)$$

and

$$\sigma = \sqrt{\frac{J^2}{n - 2}}$$

with

$$J^2 = \sum_i [d_i - a - bx_i]^2 \quad (9.61)$$

$$\Delta = \sum_i \frac{1}{\sigma_i^2} \sum_i \frac{x_i^2}{\sigma_i^2} - \sum_i \frac{x_i}{\sigma_i^2} \sum_i \frac{x_i}{\sigma_i^2} \quad (9.62)$$

9.5.1.3 The Optimization Problem

If the system can be formulated in a matrix form, the solution is the (pseudo)-inversion of the matrix.

Since the system cannot be formulated in an explicit matrix form a functional of error must be written and optimization/minimization methods should be used. As it is shown in Chap. 1, the solution is obtained comparing a number of experimental data (orders N) measured in several points of the interferometric pattern, with the theoretical predicted N values at the same points, derived from the forward model, and minimizing the difference that is a function of theoretical unknown parameters K :

$$J(K) = \sum_i^n [y_i^{pred}(K) - y_i^{meas}]^2 = \min \quad (9.63)$$

where: $y_i^{pred}(K)$ is the theoretically predicted output.

y_i^{meas} for $i = 1 \dots n$ are the experimental data. Thus:

$$J = \| N^{pred} - N^{meas} \|^2 \tag{9.64}$$

We can develop this equation, introducing the expression of N_{pred} :

$$J(K) = \sum_i^n \left[\frac{d}{f_\sigma \cdot \sqrt{2\pi}} \frac{K \cdot \sin \theta_i}{\sqrt{r_i}} - N(r_i, \theta_i) \right]^2 = \min \tag{9.65}$$

The value that minimizes J is the best estimation of K . The Least Squares method for this linear problem yields the normal equation:

$$\begin{aligned} \frac{dJ}{dK} &= 0 \\ \frac{dJ}{dK} &= \frac{K}{f_\sigma} \frac{d}{\sqrt{2\pi}} \sum_i \frac{\sin^2 \phi_i}{r_i} - \sum_i \frac{N_i \cdot \sin \phi_i}{\sqrt{r_i}} = 0 \end{aligned}$$

from which:

$$\frac{K}{f_\sigma} \frac{d}{\sqrt{2\pi}} = \frac{\sum_i \frac{N_i \cdot \sin \phi_i}{\sqrt{r_i}}}{\sum_i \frac{\sin^2 \phi_i}{r_i}}$$

All the following exercises utilize the algorithm *FindFit Mathematica*®.

Exercise 9.2 (*SIF determination for a single circular crack in an uniform state of stress*) The constant C is:

$$C = \frac{d}{f_\sigma \sqrt{2\pi}} = \frac{16}{0.688} = 23.256$$

The model compared to experimental data is:

$$C \cdot \frac{K \cdot \sin \phi_i}{\sqrt{r_i}} \rightarrow N(r_i, \phi_i)$$

This fits the experimental values with the Eq. 9.79.

The advantage of an optimization technique is a comparison between theoretical and experimental K values at several points distributed on a large portion of the definition domain, unlike the previous extrapolation of data in a single direction. Versus a theoretical value of $0.825 \text{ MPa} \sqrt{\text{mm}}$ this procedure gives an experimental evaluation equal to $0.812 \text{ MPa} \sqrt{\text{mm}}$, Fig. 9.14, that is a good approximation, versus the previous one in a direction extrapolated value of only $0.688 \text{ MPa} \sqrt{\text{mm}}$.

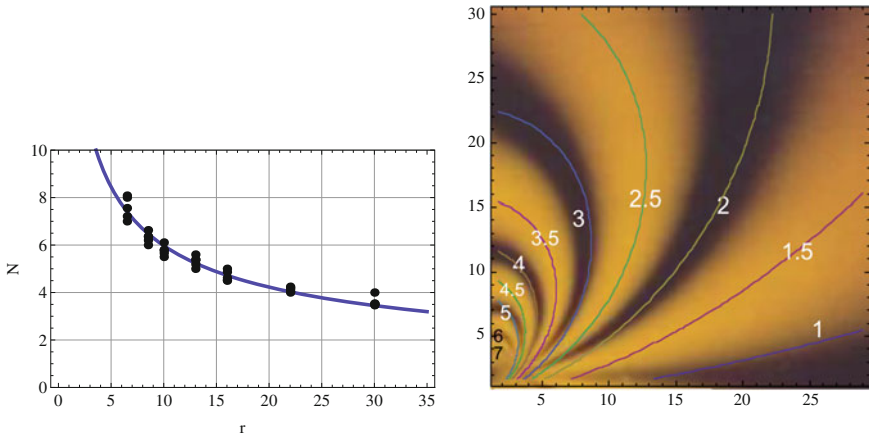


Fig. 9.14 Fitting the experimental data with a theoretical model along r and in the whole field. The K value found is $0.8120\text{MPa}\sqrt{\text{mm}}$

9.5.1.4 Non-coplanar Defects in a Thick Body

An experiment was performed for determining the stress intensity factors (in the direction of minimum distance), of two interacting circular cracks located in parallel planes of a massive body, during a stop of the fatigue propagation phase due to cyclically variable internal pressure, Fig. 9.15. Also in this case, a physical model of a homogeneous and isotropic and transparent material had been chosen, in order to visualize the front tips directions at several propagation stages and, for a material suitable for a three dimensional photoelastic analysis, to freeze the stress state in selected configurations. Figure 9.16 shows the visualization of the front shapes of

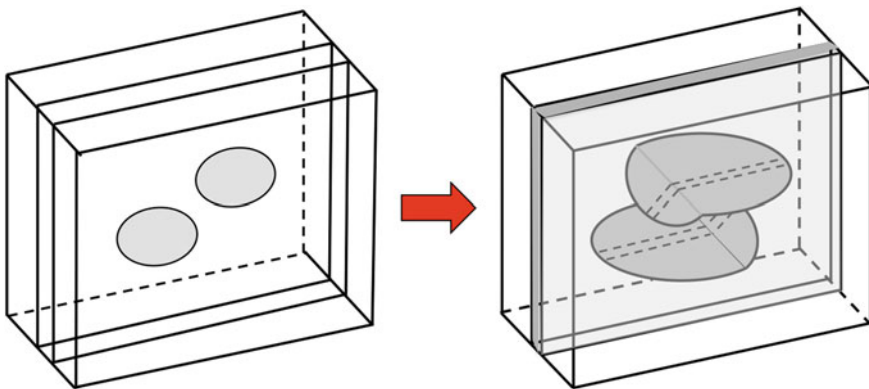


Fig. 9.15 Fatigue propagation of non-coplanar defects, initially of *circular shape*, internally pressurized

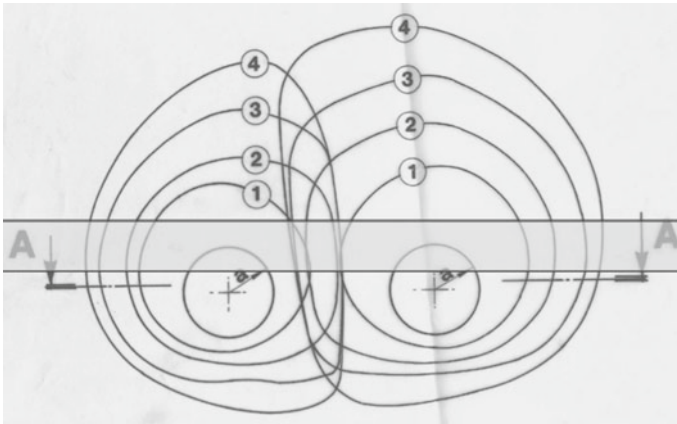


Fig. 9.16 Front shapes of non-coplanar defects for an increasing number of fatigue cycles and the perpendicular slice profile A–A

no-coplanar defects, when both are internally cyclically pressurized at increasing number of cycles. A perpendicular slice, whose profile A–A is shown, is the slice that is selected for the photoelastic analysis at one propagation stage of both cracks. This slice is cut after a freezing procedure, described in Chap. 3, Sect. 3.12.2.

A question arises as to whether this experiment can be utilized for studying the Stress Intensity Factor for two interacting cracks externally loaded, instead of internally pressurized.

The equivalence is not valid in general. It can be considered only qualitatively useful, with a restriction to the mean plane A–A in which the state of stress is essentially two-dimensional. The original application of this case was of different nature [22], but it can be an interesting case study.

Figure 9.17 shows the photoelastic fringe pattern around the tips of two interacting non-coplanar cracks, internally pressurized, in a large block of isotropic material.

A Non-linear Identification Problem

Equation 9.64 must be applied to the two cracks geometry. Unlike the case of Eq. 9.65, in the case of two non-coplanar defects, there are two variables to be identified, K_I and K_{II} , and the equation is no longer linear with respect to these variables [23], now being:

$$\left[\frac{f_\sigma}{d} N(r_i, \phi_i) \right]^2 = \frac{1}{2\pi r_i} \left[\left(K_I \sin \phi_i + 2K_{II} \cos \phi_i \right)^2 + \left(K_{II} \sin \phi_i \right)^2 \right] \quad (9.66)$$

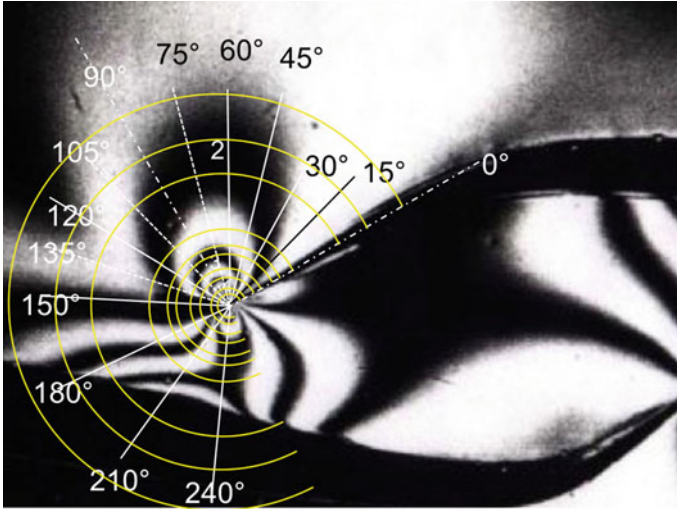


Fig. 9.17 Frozen fringes in the mean plane of two non-coplanar interacting cracks located in a large block, loaded by internal pressure

The minimization is possible through the application of an *Incremental/Iterative* approach of the Least-Squares Method.

The error functional in two variables, starting from the previous equation, is now:

$$\begin{aligned}
 J &= \sum_i^n \left\{ [(K_I \sin \phi_i + 2K_{II} \cos \phi_i)^2 + (K_{II} \sin \phi_i)^2] - \left(\frac{f_\sigma}{d} \right)^2 2\pi r_i N^2(r_i, \phi_i) \right\}^2 \\
 &= \min
 \end{aligned}
 \tag{9.67}$$

where N is the measured orders and i is the i -point label, variable from 1 to n , total number of measurements. The first term of the sum is the i th theoretically *predicted* (P) value while the second term is proportional to the fringe order *measured* (M) at the i th point. The sum is extended to all the n points.

Approximation by Linearization

The previous non-linear expression is linearized by a Taylor series expansion, limited to the first term:

$$\begin{aligned}
 P(K_I, K_{II}, \phi_i) &= [(K_I \sin \phi_i + 2K_{II} \cos \phi_i)^2 + (K_{II} \sin \phi_i)^2] \\
 &\approx P_0^2(r_i, \phi_i, K_{I0}, K_{II0}) + \left(\frac{\partial P^2}{\partial K_{I0}} \Big|_0 \Delta K_I + \frac{\partial P^2}{\partial K_{II0}} \Big|_0 \Delta K_{II} \right)
 \end{aligned}$$

Substituting the first term of a Taylor series into the error function, the problem is reduced to the minimization of a linear functional:

$$\begin{aligned}
 J &= \sum_i^n \left\{ \left(\frac{f_\sigma}{d} \right)^2 2\pi r_i N^2(r_i, \phi_i) - P_0^2 - \left(\frac{\partial P^2}{\partial K_{I0}} \Big|_0 \Delta K_I + \frac{\partial P^2}{\partial K_{II0}} \Big|_0 \Delta K_{II} \right) \right\}^2 \\
 &= \min
 \end{aligned}
 \tag{9.68}$$

Linear Least-Squares Newton-Raphson Method

In order to apply the *Linear Least-Squares Newton-Raphson Method*, minimization goes on with the corresponding normal equations in the variables:

$$\Delta K_I, \Delta K_{II}$$

At the minimum value, the derivative vector with respect to both parameters has its components equal to zero:

$$\begin{aligned}
 \frac{\partial J}{\partial K_I} &= 2 \sum_i \left[\dots \right] \frac{\partial P^2}{\partial K_I} = 0 \\
 \frac{\partial J}{\partial K_{II}} &= 2 \sum_i \left[\dots \right] \frac{\partial P^2}{\partial K_{II}} = 0
 \end{aligned}
 \tag{9.69}$$

thus:

$$\begin{aligned}
 \frac{\partial J}{\partial K_I} &= \sum_i \left[\left(\frac{f_\sigma}{d} \cdot N(r_i, \phi_i)^2 - P_0^2 \right) \cdot \frac{\partial P^2}{\partial K_I} - \left(\frac{\partial P^2}{\partial K_I} \partial K_I + \frac{\partial P^2}{\partial K_{II}} \partial K_{II} \right) \cdot \frac{\partial P^2}{\partial K_I} \right] \\
 &= 0
 \end{aligned}$$

For the first vector:

$$\begin{aligned}
 &\sum_i \left(\frac{\partial P^2}{\partial K_I} \cdot \frac{\partial P^2}{\partial K_I} \right) \partial K_I + \sum_i \left(\frac{\partial P^2}{\partial K_{II}} \cdot \frac{\partial P^2}{\partial K_I} \right) \partial K_{II} \\
 &= \sum_i \left(\left(f_\sigma/d\sqrt{2\pi r} \right)^2 \cdot N(r_i, \phi_i)^2 - P_0^2 \right) \cdot \frac{\partial P^2}{\partial K_I}
 \end{aligned}$$

For the second vector:

$$\sum_i \left(\frac{\partial P^2}{\partial K_I} \cdot \frac{\partial P^2}{\partial K_{II}} \right) \partial K_I + \sum_i \left(\frac{\partial P^2}{\partial K_{II}} \cdot \frac{\partial P^2}{\partial K_{II}} \right) \partial K_{II}$$

$$= \sum_i \left((f_\sigma / d\sqrt{2\pi r})^2 \cdot N(r_i, \phi_i)^2 - P_0^2 \right) \cdot \frac{\partial P^2}{\partial K_{II}}$$

In matrix form, for finite differences ΔK_I and ΔK_{II} (remembering that all the partial derivatives as well as P_0^2 are functions of K_I , K_{II} and ϕ_i), the following relationships shall be valid, see [21]:

$$\begin{aligned} & \begin{pmatrix} \sum_i \left(\frac{\partial P^2}{\partial K_I} \cdot \frac{\partial P^2}{\partial K_I} \right) & \sum_i \left(\frac{\partial P^2}{\partial K_{II}} \cdot \frac{\partial P^2}{\partial K_I} \right) \\ \sum_i \left(\frac{\partial P^2}{\partial K_I} \cdot \frac{\partial P^2}{\partial K_{II}} \right) & \sum_i \left(\frac{\partial P^2}{\partial K_{II}} \cdot \frac{\partial P^2}{\partial K_{II}} \right) \end{pmatrix} \cdot \begin{pmatrix} \Delta K_I \\ \Delta K_{II} \end{pmatrix} \\ &= \begin{pmatrix} \sum_i \left((f_\sigma / d\sqrt{2\pi r_i})^2 \cdot N(r_i, \phi_i)^2 - P_0^2 \right) \cdot \frac{\partial P^2}{\partial K_I} \\ \sum_i \left((f_\sigma / d\sqrt{2\pi r_i})^2 \cdot N(r_i, \phi_i)^2 - P_0^2 \right) \cdot \frac{\partial P^2}{\partial K_{II}} \end{pmatrix} \end{aligned}$$

Briefly:

$$\mathbf{A} \cdot \Delta \mathbf{K} = \mathbf{B} \quad (9.70)$$

where \mathbf{A} and \mathbf{B} are two (2×2) -matrices and $\Delta \mathbf{K}$ is a (2×1) -vector.

Partial derivatives have the following expressions:

$$\frac{\partial P^2}{\partial K_I} = 2K_I \sin^2 \phi_i + 4K_{II} \sin \phi_i \cos \phi_i$$

$$\frac{\partial P^2}{\partial K_{II}} = 8K_{II} \cos^2 \phi_i + 4K_I \sin \phi_i \cos \phi_i + 2K_{II} \sin^2 \phi_i$$

The minimization procedure starts with a guess for K_{I0} and K_{II0} and computes the values: $P_0^2(r_i, \phi_i, K_{I0}, K_{II0})$. Equations 9.70 must be solved for ΔK_I ΔK_{II} . The tentative values K_{I0} , K_{II0} must be incremented by this amount and the process must be repeated incrementally and iteratively until a minimum value for J is reached.

Exercise 9.3 (*Least-squares Newton-Raphson Method for K_I , K_{II} identification*)

- Internal pressure in both cracks: $p = 0.0866$ MPa.
- External stress applied in longitudinal direction: $\sigma_{0x} = 0$.
- Number of fatigue cycles (0 to 0.0866) of internal pressure $N_{cycl} = 32,400$.
- Starting tentative values, Fig. 9.17:
 $K_I = 0.825$ MPa $\sqrt{\text{mm}}$
 $K_{II} = 0.157$ MPa $\sqrt{\text{mm}}$.

Solution

- A minimum of $J (= 10.76)$, is reached after about 27 iterations.
- The final values are: $K_I = 0.739$ MPa $\sqrt{\text{mm}}$ and $K_{II} = 0.283$ MPa $\sqrt{\text{mm}}$.

9.5.2 Identification of SIF by Holographic Interferometry

Holographic interferometry can also be successfully applied to the determination of Fracture Mechanics parameters, remembering the relationship that links the displacements field.

9.5.2.1 Two Circular Co-planar Cracks Interaction

The application deals with the problem of two circular cracks lying on the same plane, uniformly loaded in a direction perpendicular to the plane. The experiment is set up through two cracks that are artificially generated in the middle plane of a transparent plexiglass block, Fig. 4.9 and Eq. 4.13 in Chap. 4.

The experiment wishes to investigate the propagation effect for cycling load and verify if cracks tend to join, forming a single more dangerous crack. This intuitive conjecture needs a physical validation.

From the point of view of Fracture Mechanics the stress state at any points of cracks tips is equivalent to the state due to external uniform load for *plane strain condition*.

The displacements contour lines for a given internal pressure $p = 2$ bar are shown in Fig. 9.18. The small shift of the centre of the circular contour lines with respect to the geometrical centre of the cracks is pointed out by a small eccentricity of the contour lines. Their distances from the cracks border in direction A are a little bigger

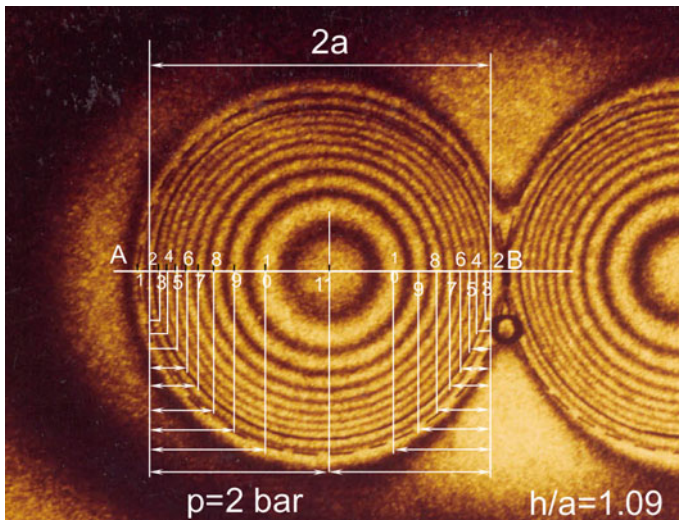


Fig. 9.18 Detail of the contour lines of the displacements field of the circular surfaces of two co-planar cracks, artificially produced in a transparent model. Uniform state of stress due to internal pressurization of $p = 2$ bar

than the distances in direction B . This small displacement is the symptom of a bigger value of the stress intensity factor in direction B with respect to direction A , as it will be proved below.

For the determination of the Stress Intensity Factor in the two radial directions at points A and B , the method of optimization of the error functional, based on the Least Squares Solution, was utilized. Remembering the equations⁷:

$$\begin{cases} u = \frac{K_I}{G} \sqrt{\frac{r}{2\pi}} \cos \frac{\theta}{2} \left(1 - 2\nu + \sin^2 \frac{\theta}{2}\right) \\ d = \frac{K_I}{G} \sqrt{\frac{r}{2\pi}} \sin \frac{\theta}{2} \left(2 - 2\nu - \cos^2 \frac{\theta}{2}\right) \end{cases} \quad (9.71)$$

a solution can be found minimizing the object function:

$$J = \|d^{predict} - d^{meas}\|^2 = \min \quad (9.72)$$

i.e. minimizing the norm of the vector of the residuals.

Substituting the second equation of Eq. 9.71 in Eq. 9.72, and also for Eq. 4.13 the following relationship is derived:

$$J = \sum_i^n \left[\frac{K_I}{G\sqrt{2\pi}} \sqrt{r_i} \sin \frac{\theta_i}{2} \left(2 - 2\nu - \cos^2 \frac{\theta_i}{2}\right) - \frac{\lambda}{2 \cos \gamma \cos \delta} N(r_i) \right]^2 = \min$$

where, with reference to Figs. 4.9b and 4.10 in Sect. 4.6 in Chap. 4:

- $G = E/2(1 + \nu)$ = shear modulus of elasticity
- ν = Poisson's ratio
- r_i = distance from the crack tip, along direction θ_i
- λ = wave length of laser light
- $\cos \gamma \cos \delta$ = geometrical position of the model on the holographic bench.

For $\theta = \pi$ the functional Eq. 9.72 becomes:

$$J = \sum_i^n \left[K_I \sqrt{r_i} - \frac{\lambda G}{2(1 - \nu) \cos \gamma \cos \delta} \sqrt{\frac{\pi}{2}} N(r_i) \right]^2 = \min \quad (9.73)$$

where:

- $K_I \sqrt{r_i} = N^{predict}$
- $N(r_i) = N^{meas}$
- $\frac{\lambda G}{2(1 - \nu) \cos \gamma \cos \delta} \sqrt{\frac{\pi}{2}} = Const.$

⁷ In the following formulas the displacement ν is called d .

Normal Equations

The corresponding normal equation is:

$$\frac{\partial J}{\partial K_I} = \frac{K_I}{G\sqrt{2\pi}} \sum_i^n \left[\sqrt{r_i} \sin \frac{\theta_i}{2} \left(2 - 2\nu - \cos^2 \frac{\theta_i}{2} \right) \right]^2 - \sum_i^n \frac{N(r_i)\lambda}{2 \cos \gamma \cos \delta} \left[\sqrt{r_i} \sin \frac{\theta_i}{2} \left(2 - 2\nu - \cos^2 \frac{\theta_i}{2} \right) \right] = 0$$

The Least Squares Solution is equal to:

$$K = \frac{\lambda G}{\cos \gamma \cos \delta} \sqrt{\frac{\pi}{2}} \cdot \frac{\sum_i^n N_i \left[\sqrt{r_i} \sin \frac{\theta_i}{2} \left(2 - 2\nu - \cos^2 \frac{\theta_i}{2} \right) \right]}{\sum_i^n \left[\sqrt{r_i} \sin \frac{\theta_i}{2} \left(2 - 2\nu - \cos^2 \frac{\theta_i}{2} \right) \right]^2} \tag{9.74}$$

By the *LSM*, *K* is obtained in directions *A* and *B* for a loading pressure ($p = 2$). The result of the analysis is $K_B/K_A = 1.08$.

Exercise 9.4 (*The application of the LSM by a direct algorithm, (avoiding writing the normal equation)*) The application of the *FindFit* algorithm simplifies the calculation of normal equations because it directly provides the expression of K_I in points *A* and *B*. The algorithm works directly on the Eq. 9.73. Figure 9.19 shows the two curves that fit to the experimental data, minimizing the residuals, providing both values of the stress intensity factors. For an internal pressure of $p = 2$ bar the values are the following:

- K in Direction A $K = 1.261 \text{ MPa} \sqrt{\text{mm}}$
- K in Direction B $K = 1.362 \text{ MPa} \sqrt{\text{mm}}$
- $K_B/K_A = 1.080$

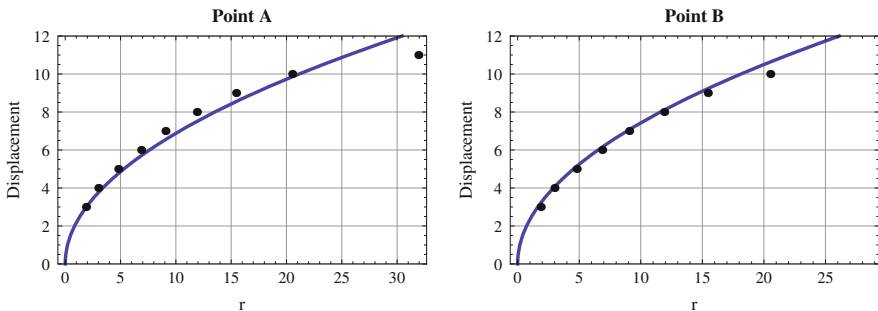


Fig. 9.19 The two curves of the algorithm *FindFit* in Mathematica® for fitting holographic interferometry data in the case of $p = 2$ bar. The constant is 0.5800

The value derived from LSM (Least Squares Method) application for $p = 3$ bar is $K_B/K_A = 1.102$.

The holographic interferometry images prefigure the propagation shape of both cracks, showing a quasi-elliptic zone of influence to which the defects have a tendency after junction, Fig. 4.12 in Chap. 4.

The magnification effect $M = K_B/K_A$ in the direction of minimum distance between the cracks is about $10 \div 11\%$ greater than that of a single crack and is much more independent than the single values K_A and K_B , from the material parameters and loadings (E , ν , p , etc.). This is consistent with Chap. 1, Sect. 1.1 which recommends selecting the unknown output parameters so that they are insensitive to other model parameters selected in the experiment, as it happens in this case where the material of the model (Plexiglass) is different from the structural material in which the real cracks arise, but it does not influence this result.

The industrial application that suggested this experiment is related to the propagation of very small but numerous circular cracks located in rotors of turbomachinery. The small circular defects formation occurs in the practice of the industrial production of molding and subsequent forging phases of large steel ingots. The propagation arises by centrifugal forces acting in low cycle fatigue stage, Chap. 8, due to starts and stops of the rotors in normal functioning conditions.

Ultrasonic investigations are able to detect defects that are presumed, in the absence of other information, to be circular and lying in the most stressed radial planes. The small co-planar cracks, if close to one another, can interact, i.e. magnify the stress intensity factors, with the risk of a possible junction, after a sufficient number of fatigue cycles.

In the present example of two co-planar defects cyclically pressurized, the evolution of their shapes is shown in Figs. 9.39 and 9.40. The fronts evolve into a common quasi-elliptical shape that later becomes circular.

9.5.3 Identification of the Stress Intensity Factor by Strain Gages

Strain gages measurements offer several ways to identify the plane-stress K value, with one or more higher order terms of series expansion of strain close to the crack tip. Let us assume that strain gages can be placed in a region near (but not very near) the crack tip along the crack axis, with their grids in the direction perpendicular to the crack plane [13]. The case study deals with a corner crack in a component of rectangular section in a portal frame consisting of two columns, subjected to tensile stress, and two crossheads.

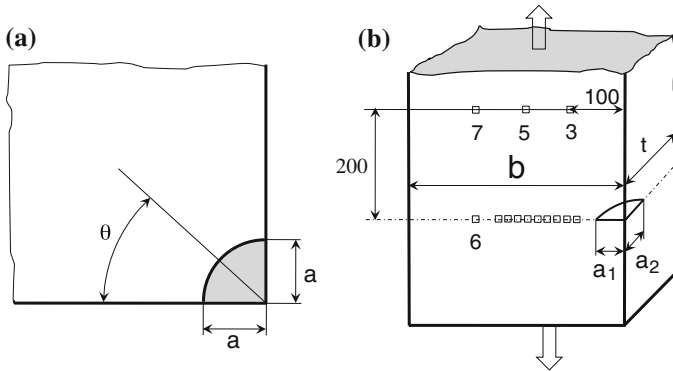


Fig. 9.20 Sketch of angular crack in a tensile member. In **a** a theoretical form of a quarter of circular crack in infinite body; in **b** the real shape of angular crack after propagation

During the tests on a prototype, after a relevant number of working cycles, an angular crack enucleated in a corner that developed in the form of a quarter of a circle and then in a more complex shape, Fig. 9.20a. Data are the following:

- Material: Fe 510 with $E = 2.1 \cdot 10^5$ MPa
- $\nu = 0.3$ Poisson's ratio
- $b > 400$ mm column width
- $t > 600$ mm column thickness
- $a_1 \geq 45$ mm crack length for $3.5 \cdot 10^6$ cycles
- $a_2 \geq 45$ mm that grows up to 200 mm for $3.5 \cdot 10^6$ cycles.

A chain of nine strain gages is located on the crack axis on the shorter side of the crack, at a distance $15 \div 30$ mm from the crack tip, Fig. 9.20b.

According to the model of linear elastic fracture mechanics (LEFM), three regions are defined surrounding the crack tip, see Sect. 9.3. Locating the strain gages either in a zone very close to the tip where the material yields with large deformation, or in a too far region, where boundary effects dominates and the model is no longer representative of the state of the local stress, must be avoided.

The measured values of the nine strain gages along the crack plane are reported in Table 9.1.

9.5.3.1 Identification by the First Term of the Series

Measurements of strains along the crack plane can offer enough information for the K_I identification. In order to fit the theoretical model to the data, the Least Squares method is applied as follows, comparing the predicted strains with the measured ones:

$$J = \|\varepsilon^{predict} - \varepsilon^{meas}\|^2 = min \tag{9.75}$$

Table 9.1 Strain measurements along the crack plane

# SG	Distance r_i (mm)	Meas. strain ($\mu\epsilon$)
1	15.30	762
2	16.90	699
3	19.15	656
4	20.80	623
5	22.80	591
6	24.70	566
7	26.90	557
8	29.10	537
9	30.70	527

$$J = \sum_i^n \left[\frac{(1 - \nu)}{E\sqrt{2\pi}} \cdot \frac{K_I}{\sqrt{r_i}} - \epsilon^{meas}(r_i) \right]^2 = \min \tag{9.76}$$

where:

$$\epsilon^{predict} = \frac{(1 - \nu)}{E\sqrt{2\pi}} \cdot \frac{K_I}{\sqrt{r_i}} \tag{9.77}$$

Exercise 9.5 (Identification of K_I by the First term of the Series Expansion with the previous data) The constant C is equal to:

$$C = \frac{(1 - \nu)}{E\sqrt{2\pi}} = \frac{0.7}{2.507 \cdot 2.1 \cdot 10^5} = 1.330 \cdot 10^{-6} \frac{1}{\text{MPa}}$$

By *FindFit* of *Mathematica*® [24], the identified value of the stress intensity factor with only the first term of series expansion is: $K_I = 2, 166.85 \text{ MPa}\sqrt{\text{mm}}$, Fig. 9.21a.

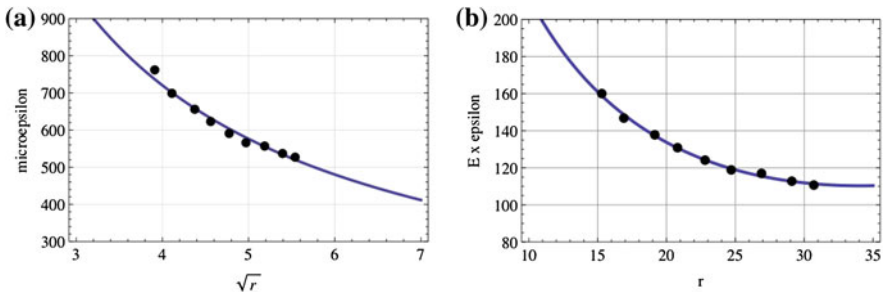


Fig. 9.21 Theoretical model fitting experimental data according to Eq. 9.76. **a** With one term, and **b** with three terms of the series expansion. Note the difference of the variables on the horizontal axis: in **a** is \sqrt{r} and in **b** is r

9.5.3.2 Identification by More Terms of the Strains Series Expansion

In order to check the accuracy of K identification by strain gages, the use of more terms of the series expansion is suggested because the chain of strain gages, due to their natural transversal dimensions, tends to overcome the limited area close to the crack tip and they could record strain contributions at points relatively far from the tip to justify the use of models with more terms (almost four) of the series.

The relationships for the strains can be derived substituting Eqs. 9.41 and 9.42 into the general stress-strain relations (Hooke's law); for the strain component ε_y perpendicular to the crack line, the equation is:

$$\begin{aligned}
 E\varepsilon_y = & A_0 r^{-1/2} \cos \frac{\theta}{2} \left[(1 - \nu) + (1 + \nu) \sin \frac{\theta}{2} \sin \frac{3\theta}{2} \right] - 2\nu B_0 \\
 & + A_1 r^{1/2} \cos \frac{\theta}{2} \left[(1 - \nu) - (1 + \nu) \sin^2 \frac{\theta}{2} \right] \\
 & + A_2 r^{3/2} \left[(1 - \nu) \cos \frac{3\theta}{2} + \frac{3}{2} (1 + \nu) \sin \theta \sin \frac{\theta}{2} \right]
 \end{aligned} \tag{9.78}$$

For $\theta = 0$. Eq. 9.78 is reduced to the following:

$$E\varepsilon_y = A_0 r^{-1/2} (1 - \nu) - 2\nu B_0 + A_1 r^{1/2} (1 - \nu) + A_2 r^{3/2} (1 - \nu) \tag{9.79}$$

Exercise 9.6 (Fit the experimental data of Table 9.1 by Eq. 9.79, with three (A_0 , A_1 , A_2) terms of a series expansion and compare the result of this interpolation with the corresponding result obtained with only one term) Through the use of algorithm *FindFit*, the estimated value of the Stress Intensity Factor is $K_I = 2973.38 \text{ MPa} \sqrt{\text{mm}}$, Fig. 9.21b.

The improvement of the three terms model observing the differences between the diagrams a and b of Fig. 9.21 seems minimum, but the estimation of the Stress Intensity Factor changes considerably: 2973.4 instead of 2166.8 $\text{MPa} \sqrt{\text{mm}}$.

9.6 Stress Analysis for Brittle Fracture

In order to proceed with the theory of the Fracture Mechanics it is necessary to evaluate the material answer to a singularity in the state of stress, described by the K_I factor of a crack in a loaded structural element, i.e. the material capacity to tolerate the presence of a crack without fracturing. This is called *Fracture Toughness* and represents the critical value of the stress intensity factor K_{Ic} , when the specimen exhibits a sudden crack propagation.

While the corresponding behavior of the stress/strain curve in the static or in the fatigue cases was simulated by mathematical models, such as the Ramberg-Osgood equation, and by the mathematical description of the fatigue curve as well as of the hysteresis cycles in the low cycle fatigue, it is not equally possible to describe the behavior of a crack when it is loaded monotonically up to the brittle propagation of the crack. The phenomenon is much more complex and it is necessary, in this case:

1. to develop special specimens that must be representative of the real cracks in a structure,
2. to develop special testing machines suitable to realize plane state of strain. Only in this case does the test gives a result that is independent of the material thickness, since the size of the plastic zone accompanying the crack tip is very small relative to the specimen thickness, and the fracture toughness does not depend on it. As it will be shown below, this condition could require testing machines of high capacities.
3. to develop special testing protocols for covering a variety of materials behaviors.

The protocol is regulated by standards, covering the Test Methods for Linear-Elastic Plane-Strain Fracture Toughness K_{Ic} of the materials.⁸

9.6.1 Experimental Equipments

As regards the first point, the choice of the specimen type, in order to remove any doubt on the opportunity of considering the study and determination of the fracture toughness as a task of stress analysis, it must be remembered that a critical value of the stress intensity factor (i.e. the Fracture Toughness) is not only a material property but also depends on the stress state in the specimen, chosen to represent the real situation of risk of an artifact.

The Stress Intensity Factor is defined only in the elastic field, with only a small area of plasticization close to the tip: then testing must be rigorously performed in this condition, if we want to maintain similarity with the real risk of brittle fracture (i.e. a sudden fracture not announced nor followed by any visible plasticity of the structural element).

For example in Mode I, that is the most important fracture mode, since the stress near the crack tip and K_I factor are proportional one to the other, ($\sigma \cdot \sqrt{2\pi r} = K_I$) an increment of stress corresponds to an increment of the stress intensity factor, thus a crack propagation will happen when K_I reaches a critical value K_{Ic} .

⁸ For e.g.: ISO 12135: Metallic materials Unified method of test for the determination of quasi-static fracture toughness, First edition 2002-12-01 with a Technical Corrigendum 1, Published on 2008-06-01.

9.6.1.1 Specimens

The classic test procedure needs a proper choice of a cracked specimen. The most common forms suggested by international previously quoted ISO Standards are the Compact Test Specimen *CT*, Fig.9.22 and the three-points bending specimen, prepared with a sharp notch exacerbated with a pre-cracking operation (in order to simulate real cracks), obtained by a fatigue controlled test.

The calibration formula for K_I is approximately obtained by a series expansion as a function of the non dimensional term a/W which compares with powers of $1/2$; $3/2$; $5/2$; $7/2$; $9/2$.

The approximation derives from the complex boundary conditions of the specimen that was developed for reducing the quantity of material necessary for the characterization (Fig.9.23):

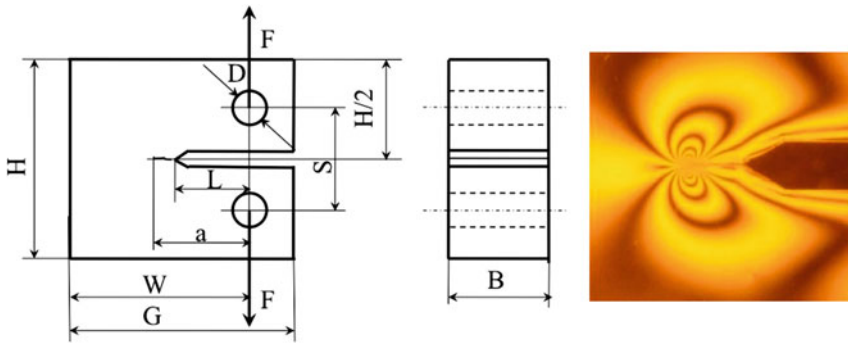
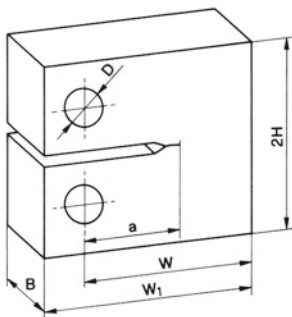


Fig. 9.22 CT specimen proportion: ($B = 1/2 W$ $a = 0.45 \div 0.55 W$ $S = 2 \cdot 0.275 W$) and the photoelastic fringes of the stress distribution around the tip



Size	B	W	a	W_1	H	D	TEST
2TCT	50	100	50	125	60	25	K_{IC}
4TCT	100	200	100	250	120	50	"
5TCT	125	250	125	312.5	150	63.5	"
6TCT	150	300	150	375	180	75	"
1/5CT	5	25	5	31	15	5	Propag.

Fig. 9.23 Typical dimensions of the *CT* Compact Test specimen for toughness measurements (length in mm), see details in [25]

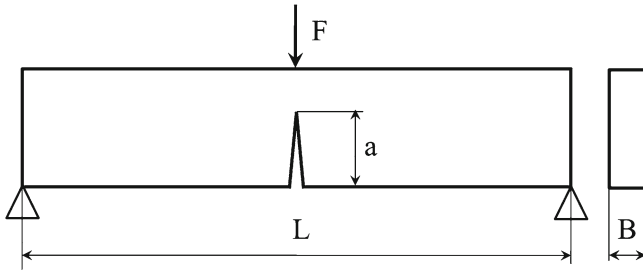


Fig. 9.24 The three-point bending specimen

$$K_I = \frac{F}{BW^{1/2}} \left[29.6 \left(\frac{a}{W} \right)^{1/2} - 185.5 \left(\frac{a}{W} \right)^{3/2} + 655.7 \left(\frac{a}{W} \right)^{5/2} - 1017 \left(\frac{a}{W} \right)^{7/2} + 638.9 \left(\frac{a}{W} \right)^{9/2} \right] \quad (9.80)$$

This equation is valid for $0.45 \leq a/W \leq 0.55$ and for the following dimensioning:

$$\begin{aligned} W &= 2B & a &= 0.55 B & S &= 0.55 W & H &= 1.2 W \\ a &= 0.45 \div 0.55 W & D &= 0.25 W & W_t (\text{Sometime called } G) &= 1.25 W \end{aligned}$$

Another common specimen geometry that is used for K_{Ic} determination is the so-called three-point bending, Fig. 9.24:

$$K_I = \frac{FL}{BW^{3/2}} \left[2.9 \left(\frac{a}{W} \right)^{1/2} - 4.6 \left(\frac{a}{W} \right)^{3/2} + 21.8 \left(\frac{a}{W} \right)^{5/2} - 37.6 \left(\frac{a}{W} \right)^{7/2} + 38.7 \left(\frac{a}{W} \right)^{9/2} \right] \quad (9.81)$$

In these Equations, all the parameters are known except the force F that must be measured at the instant of sudden crack propagation.

9.6.1.2 Special Devices

With the use of the previous specimens,⁹ the toughness testing is performed by relating the applied load with crack opening displacement Δ of the crack mouth. The relative displacement of two points located symmetrically on opposite sides

⁹ Many other kinds of specimen geometry can be used. In Fracture mechanics handbooks suggest a variety of solutions with the relative calibration formulas [10, 26, 27].

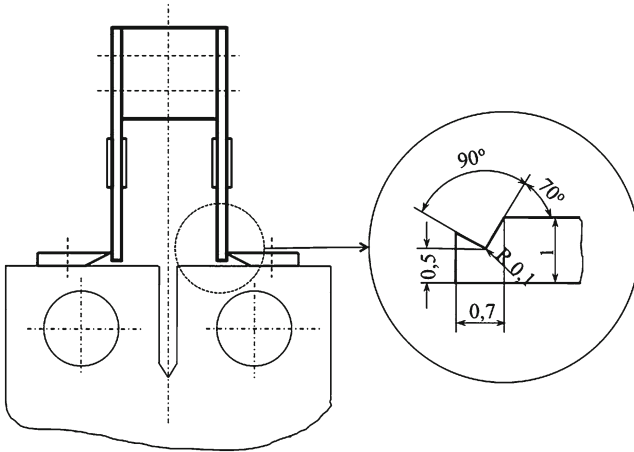


Fig. 9.25 Devices for the measurement of the crack opening displacement in the fracture toughness testing

of the crack plane of the pre-cracked specimen, is measured by a device Fig. 9.25, and recorded simultaneously with the monotonically increasing load, applied to the specimen. The specimen is loaded at a rate of stress intensity K_I/s , within the range $0.55 \div 2.75 \text{ MPa m}^{1/2}/s$.

The device has the task of measuring the distance between the two reference points and is called an *extensometer*. It is formed by two small cantilevers of harmonic steel mounted with a pre-load between the two blades, in such a way to keep the contact when the mouth enlarges.

Four strain gages are mounted on the blades in a full Wheatstone bridge and the extensometer is calibrated with a centesimal measuring tool before the test. Commercial instruments are available but a device built in-house can be adjusted at the necessary sensitivity, varying the position of the gages on the two cantilever beams.

There are other types of commercial extensometers such the one shown in Fig. 9.26.

9.6.1.3 The Specimen Thickness

The K_{Ic} parameter must be independent of the thickness of the body. In order to guarantee this condition, the specimen must be in plane strain state because only in this condition is the plastic zone size at a minimum, with a dimension on the crack plane practically equal to zero, Fig. 9.4 and with a full brittle fracture on the entire thickness. This is the most dangerous situation and this is the lower limit of the material that we wish to measure, Fig. 9.27 on the right side.

This condition occurs if the size of plastic zone is small compared with a and B :



Fig. 9.26 Commercial extensometer

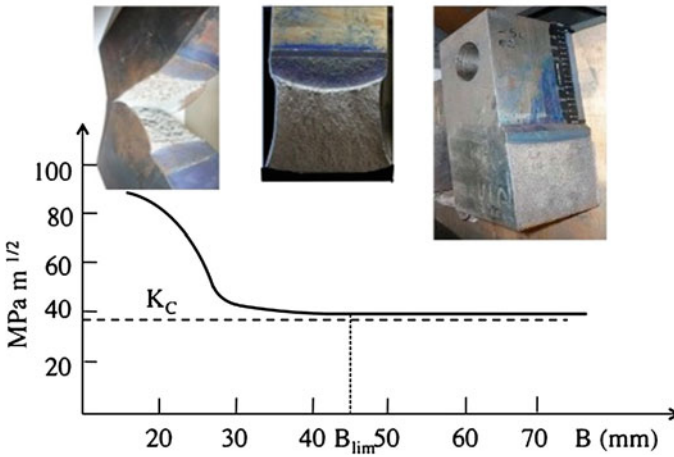


Fig. 9.27 Trend of the fracture toughness as a function of the thickness [5, 6, 25, 28]

$$r_{pl} = \frac{1}{2\pi} \cdot \left(\frac{K_I}{\sigma_{sn}} \right)^2 \leq \frac{a, B}{10 \div 20} \text{ that gives the following disequalities:}$$

$$a, B \geq 2.5 \left(\frac{K_{Ic}}{\sigma_{sn}} \right)^2 \quad e \quad \frac{a}{W} \leq 0.55 \tag{9.82}$$

where: a, B, W are the crack and body dimensions Fig. 9.22. The same Figure shows the calibration formula for calculating K_{Ic} for given dimensions and for a determined load and the *Critical Crack Length*.

The first of Eqs. 9.82 is a condition on the plastic radius r_{pl} , that must be small compared to a , while the second guarantees that a small error on a length does not generate notable variations of the value of K_{Ic} . All these conditions imply a limit on the minimum thickness of the test specimen for valid K_{Ic} measurements.

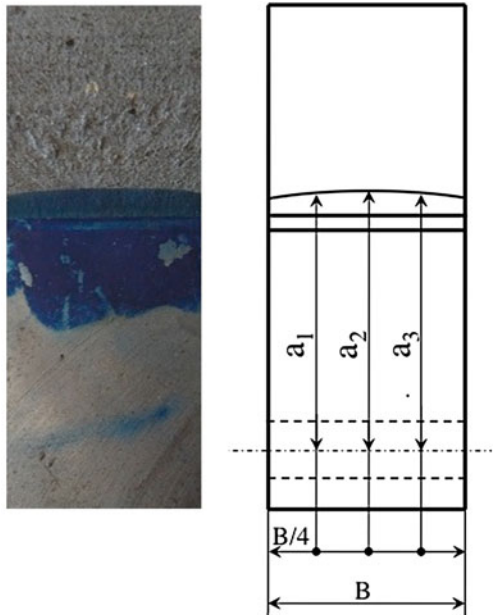
In order to choose the specimens sizes before knowing the fracture toughness of the material, it is advisable to find a first tentative value of it with a large specimen test, and to reduce the size after this first estimation. If this second value is identical to the first one, that means we are beyond the limit value B_{lim} of Fig. 9.27 and the test with the smaller thickness specimen is valid. Since the fracture toughness is strongly influenced by the temperature, the specimen thickness is dependent on it and the estimated value of K_{Ic} must be determined for the working temperature of the artifact from which the specimen derives. The estimation of the crack length is shown in Fig. 9.28.

9.6.2 Fracture Mechanics Testing

Fracture toughness determination of very tough materials requires testing apparatus in-house designed for high capacity, Fig. 9.29 in order to:

- Test thick specimens.
- Maintain the load rigorously centered on the mean plane of the specimen cross-section.
- Control the applied load that must be incremented at a constant rate.

Fig. 9.28 Estimation of the most likely value of semi-length critical crack as the average of three values



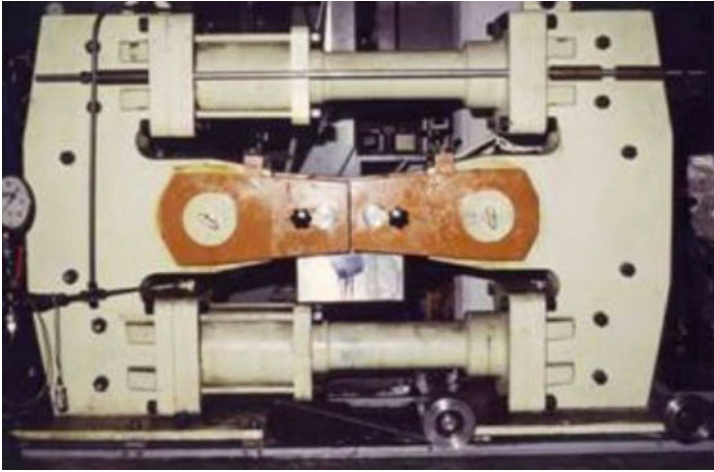


Fig. 9.29 Special self-manufactured testing apparatus [29] for the measurement of K_{IC} in compact tension specimens up to 150 mm thick, and 375×360 mm with a maximum load of 2 MN, see Fig. 9.23

- Control the *test temperature* through the use of a refrigerating fluid.

The critical load related to the value of fracture toughness is obtained from recorded curves, in agreement with the standards that are focused on assuring an estimate in real plane strain state. The procedure is stated in the following steps, Fig. 9.30:

- The mouth of the notch is instrumented by the device of Figs. 9.25 and 9.26 to detect the crack opening displacement v .
- The specimen is loaded at a constant rate F/v and the maximum force F_{max} is recorded in the diagram $F - v$.

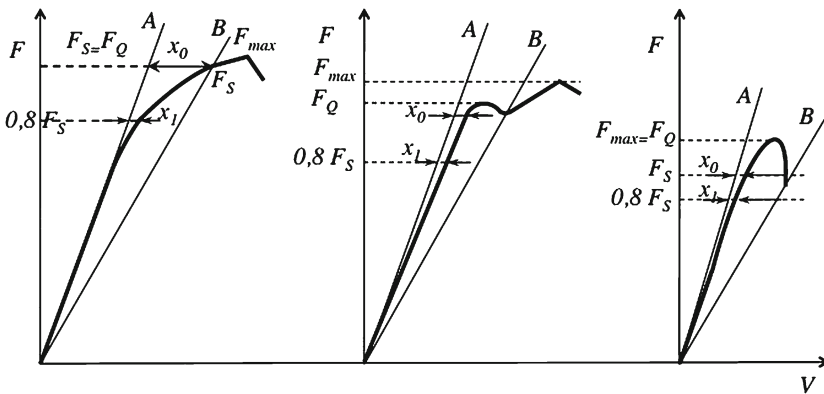


Fig. 9.30 Kinds of the curves load-crack opening at the mouth of the notch

- A secant line is drawn with a slope equal to 95 % of the slope of the straight line tangent to the first linear portion of the experimental curve: $(F/v)_{secant} = 0.95(F/v)_{tang}$.
- The ordinate of the point of intersection is stored as the value F_s .
- The ordinate value equal to $0.8 \cdot F_s$ is recorded and the segment x_1 is compared with the segment x_0 as indicated in the same Fig. 9.30.
The distance x_1 between the straight line OA and the curve, corresponding to the ordinate $0.85 F_s$ must be such that: $x_1 \leq 0.25x_0$.
- The same standard establishes that, for the proof validity, the ratio F_{max}/F_s must be less than or equal to 1.1.
If all the conditions are verified, the three possibilities shown in Fig. 9.30, present in different materials, are acceptable and the value of the force F_Q to be utilized in the calibration formulas Eqs. 9.80 or 9.81, is given respectively for the three cases, by:
 1. The value of F_s in the first case.
 2. The relative maximum of the curve lying between tangent and secant in the second case.
 3. The absolute maximum in the third case.
- The test is concluded with the measurement of the critical crack length, with the Fracture Toughness calculation and with the validity verification of the thickness of the specimen obtained by the measured value of K_{Ic} through the use of Eq. 9.82.

A Case Study on a Tentative Way for Fracture Mode I Measurement

There could be other possibilities of carrying out tests for the Fracture Toughness determination of mode I with non-conventional specimens. Even with many reservations, just to investigate a new way to find less demanding methods for Fracture Toughness determination, let us consider a cylindrical specimen of circular section loaded by pure torsion with a crack along a helical line at 45°. It can be theoretically used to achieve a state of plane strain with a small material volume, compared to that of standard specimens.

The photoelastic stress image in a section perpendicular to the helical curve in Fig. 9.31 shows that the distribution of local stresses is identical to that in a plane specimen of finite size, with a through-thickness crack, uniformly loaded. The plain strain condition is guaranteed by a very long crack and by the axial constraint along the crack that avoids contraction.

In Fig. 9.31 the stress singularity at the tip of the plane specimen is compared with the singularity at the tip of helical cracks, while Fig. 9.32 shows: (a) the specimen machining with a milling machine for obtaining the helical notch in the cylindrical specimen, (b) the fracture surface with the highlighting of the regular crack obtained in the pre-cracking phase in torsion: (c) the specimen with two rings to accommodate the measuring device of Fig. 9.23 for measuring the crack opening displacement and in (d) the curve that correlates the torque and the opening of the mouth of the crack,

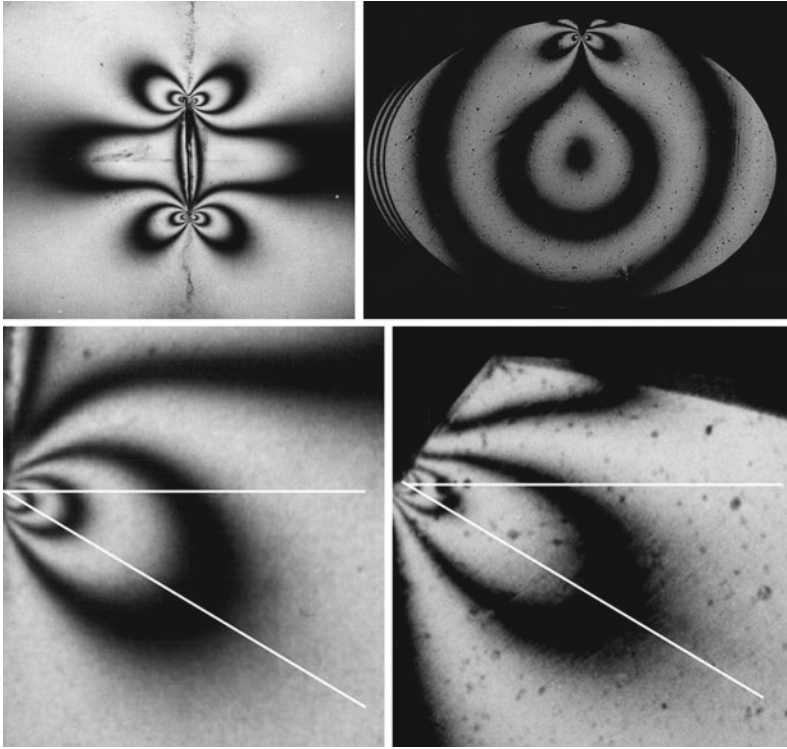


Fig. 9.31 Comparison between contour lines of the maximum shear stresses in a plane specimen and in a plane at 45° (with respect to the symmetry axis) in a cylindrical specimen in pure torsion, with a crack along a helical curve at 45° [30]

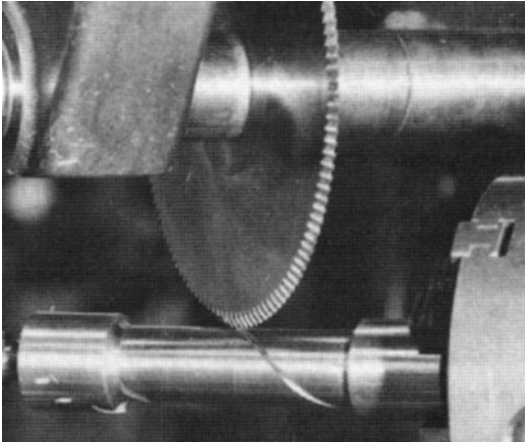
from which the value of K_{Ic} is inferred. The calibration is done identifying the Stress Intensity value by the photoelastic fringes.

For a steel of the NiCrMoV family, at the test temperature of -120°C , the average of three measurements of Fracture Toughness gave a value of $30\text{ MPa}\sqrt{\text{m}}$, consistent with the value obtained from much larger specimens of *CT* type. This method should be applied to a variety of cases and materials in order to obtain a scientific evidence.

9.6.3 A Case Study on the Temperature Influence on SIF

It is known that temperature influences the material behavior with particular effect on the material parameters connected with the strength and the fracture resistance, such as resilience and yield stress. There are mainly two kinds of fracture and a material can exhibit both types if the test temperature varies widely. Consequently it is clear that Fracture Toughness too must be strongly influenced by the temperature. This can be verified in Fig. 9.33 where Fracture Toughness and yield stress are reported as functions of temperature for a *NiCrMoV* steel together with the Fracture Appearance

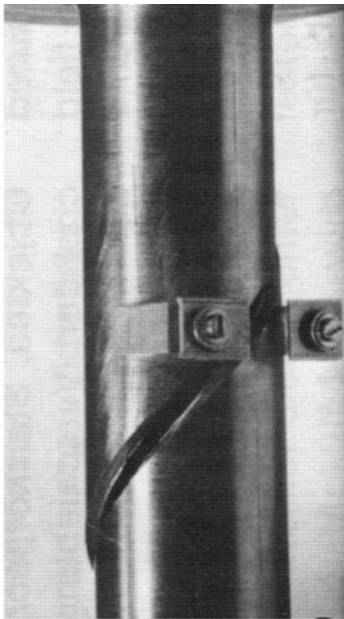
(a)



(b)



(c)



(d)

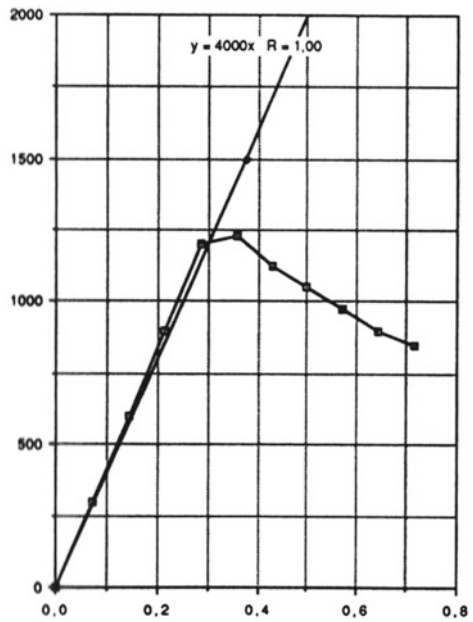


Fig. 9.32 Tentative way to perform the test characterization for the measurement of Fracture Toughness K_{Ic}

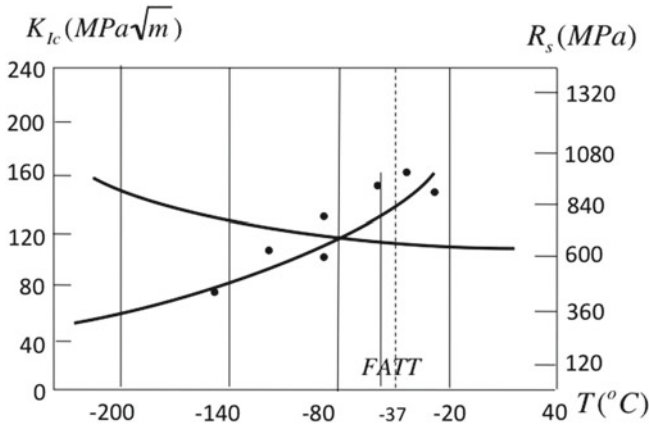


Fig. 9.33 Fracture Toughness trend (on the left) and yielding stress trend (on the right) as a function of test temperature for a NiCrMoV steel

Transition Temperature (FATT) that discriminate between the brittle $T \leq FATT$ from the ductile $T > FATT$ behavior on the basis of the aspect of the fracture surfaces.¹⁰

The influence of the test temperature on the fracture Toughness is well clarified in Fig. 9.34 for low alloy steels suitable for rotors of electrical machineries. A lowering of the temperature influences the maximum plastic zone size for a concomitant reduction of K_{Ic} and an increment of σ_Y , with a quadratic dependence. The material assumes a brittle behavior that favors unstable crack propagation.

A similar phenomenon happens for a crack progressing at high speed that has no time to realize a ductile fracture mode. The deformation rate too has thus a big influence on fracture toughness, so that a new material parameter is introduced K_{ID} a dynamic stress resistance that is lower than the static fracture toughness K_{Ic} . The high speed is generated by the releasing of stored energy of deformation that can accelerate the crack, up to reach high speeds such that it can destroy structural elements as pipelines, in few seconds.¹¹

9.7 Models for Sub-critical Fatigue Crack Growth

So far we have dealt with the analysis of cracks under static loadings and examined the critical condition of brittle fracture that occurs when the value of the stress intensity factor K_I reaches or exceeds the value of the material Fracture Toughness K_{Ic} .

¹⁰ Empirical formulas of correlation exist between this Transition Temperature and the one obtained by the Charpy test [1].

¹¹ If the speed reaches 1 km/s, 5 s are enough to destroy 5 km of a gas pipeline.

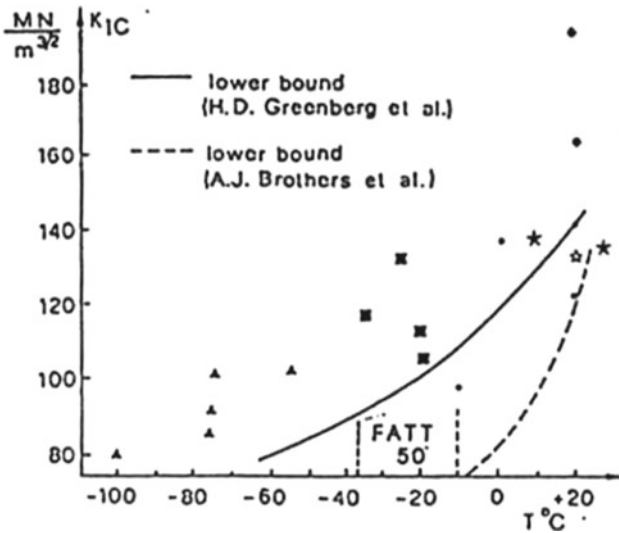
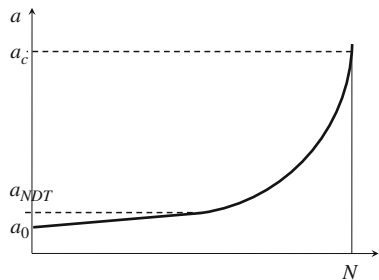


Fig. 9.34 Example of Fracture Toughness dependence on test temperature of a low alloy steel for large electrical machinery rotors (hardened and relieved *NiCrMoV* steels with $\sigma_Y = 370$ MPa), obtained by an author from compact test specimens with thickness from 25 to 150 mm [25]. Two limit curves taken from the literature have been shown: both are conservative, but especially the curve of A.J. Brothers et al. seems inadequate. Experiences performed for Ansaldo s.p.a. (Genova) [31]

Even limiting the presentation to the concepts related to the stress analysis, it is necessary to examine the phenomenology of the sub-critical crack growth when the applied load is fluctuating, both in the condition of high cycle fatigue as well of low cycle fatigue. The crack propagation reacts in different way when is affected by innocuous (inert) or aggressive environments with temperature influence.

Modeling the sub-critical crack growth from the point of view of stress state is a difficult task that has found a variety of phenomenological interpretations, expressed by a number of laws. The sub-critical crack growth for cyclic loads can be highlighted in the first instance, directly recording the advancement of crack at increasing values of the number of stress cycles. Figure 9.35 show this type of correlation, being the propagation unstable and strongly accelerating as the length of the defect increases.

Fig. 9.35 Trend of semi-length of a crack in a plate versus N the fatigue cycles. a_{NDT} is the minimum detectable value by NDT methods. a_0 is the initial semi-length



It is intuitive that, rather than finding a direct correlation between the length of the crack and the number of load cycles at a certain propagation instant, it is better to discover, at a certain instant of time, a correlation between the crack length and the value of the local stress cycle that is connected with the difference $K_{max} - K_{min} = \Delta K$.

Having observed a non linear increment of the crack length (and also of the its derivative da/dN) with the number of cycles, all the propagation laws have assumed a correlation between the range of the Stress Intensity Factor ΔK and the derivative da/dN .

Experiments have shown that in double logarithmic scale the link between da/dN and ΔK have a linear correspondence at least in the central interval of values, Fig. 9.36a. A more accurate analysis shows a weak dependence of the ratio $R = \sigma_{min}/\sigma_{max} = K_{min}/K_{max}$ that becomes significant in the low as well as in the high part of the diagram, with the trend shown in Fig. 9.36b with curves of S form. This observation, due to Paris, allows a synthesis of modeling the propagation phenomenon, because it is possible to synthesize the data for cracks of different sizes and with different loading values, simply relating the ΔK to the measurement of the rate of growth. The curves da/dN as function of $\Delta K = K_{max} - K_{min}$ for several values of $R = K_{min}/K_{max} = \sigma_{min}/\sigma_{max} =$ can be classified in three regions:

1. In the first, called region I, a characteristic value ΔK_{th} represents the threshold value of stress intensity factor behind which the rate of advancement tends to zero value and the logarithm to $-\infty$.
2. For intermediate values of K (region II), it is observed a linear law between $Log da/dN$ and $Log \Delta K$, that can be described, in the most simple way, by the Paris law that:

$$da/dN = C \cdot (\Delta K)^m \tag{9.83}$$

where C and m are material constants. The reference test is performed with a cyclic load with zero minimum value and the influence of R and load frequency is considered negligible.

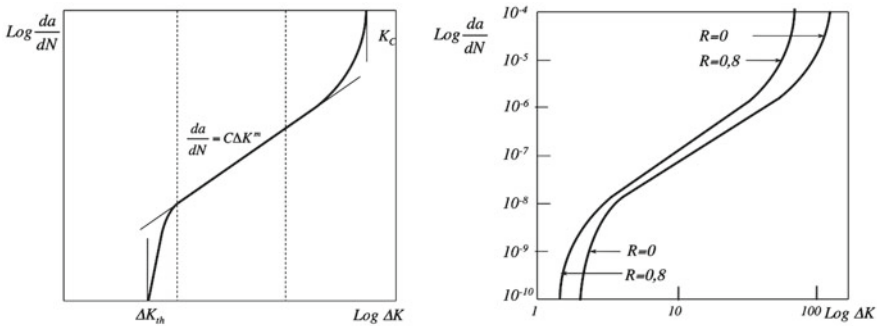


Fig. 9.36 Trend of crack rate with ΔK and influence of R

3. For high values of ΔK (region III) the behavior becomes unstable with a rapid increment of the propagation rate up to the total specimen fracture. This condition is reached for a stress intensity factor equal to a critical value a K_c , not necessarily equal to K_{Ic} since the specimen cannot be in the condition of plane strain.

Two different conditions can be verified in this sub-critical growth test:

- The first condition is realized for high resistance materials with low toughness: the specimen behavior remains elastic up to a value of the Stress Intensity factor equal to K_{Ic} . In his case the maximum stress in the loading cycle is such that, for a critical crack length, the stress intensity factor reaches the critical value K_{Ic} maintaining plane strain condition: in this case the fracture is brittle.

Other relationships have been developed for taking into account of the influence of other variables as the critical factor K_{Ic} and the ratio R . One of the most used is the *Forman* equation:

$$\frac{da}{dN} = \frac{C \cdot (\Delta K)^n}{(1 - R) \cdot (K_c - \Delta K)} \tag{9.84}$$

where C ed n are material constants, of course, other than the Paris equation. This equation describes the type of behavior for accelerated da/dN for high values of ΔK , that is not taken into account by the Paris model.

The Forman model also accounts for the influence of R and describes the behavior in the case where K is close to K_c . This occurs for metals at high resistance and low tenacity in which the specimen sizes justify elastic behavior also for values of K close to K_{Ic} .¹²

- The second condition is verified when the crack length reaches a certain value, the frontal specimen ligament yields overcoming the elastic limit, entering the plastic behavior. This happens for ductile materials. When the yield controls the phenomenon the ΔK values are no more referable to the fracture toughness. The limit of small yielding zone, hypothesized for the linear elastic fracture mechanics, is overcome.

New concepts for the elastic-plastic fracture mechanics must be introduced as the Rice integral J , or *COD* [1].

9.7.1 Case Study on Sub-critical Crack Growth for Turbine Steels and Design Criteria

The following reflexions have been suggested about the modeling of sub-critical crack growth in different case studies:

¹² Let observe the difference between K_c and K_{Ic} . Due to the different tests conditions (between the static Fracture Toughness and propagation tests), not necessarily plane strain conditions are realized in sub critical growth tests. Thus the critical value K_c is, in general, a function of the thickness B of the specimen.

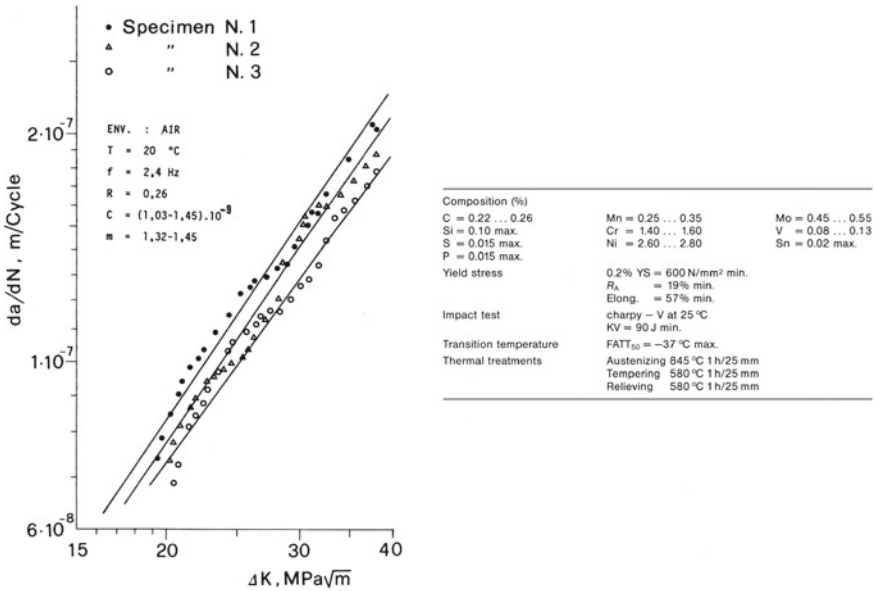


Fig. 9.37 Crack propagation data for a NiCrMoV steel in region II in an inert environment [25]

- Figure 9.37 show tests examples on NiCrMoV steels for low-pressure turbine discs with a thermal treatment of austenizing, hardening and relieving as described in [25]. Three straight lines for three specimens series show the band of dispersion for the same material and the same test conditions. As consequence the material parameters C and m are subjected to a variability that can be estimated through the Eq. 9.59. Thus it is suitable to obtain from the tests not deterministic values but mean values with the relative variances. Besides, the behavior in region II is important for components subjected to a finite number of fatigue.
- The threshold value ΔK_{th} is a *design parameter* for those applications, as low stressed rotating shafts at high cycle fatigue, for which is not considered as useful life the cycles spent in the sub-critical propagation phase. In region I the defects growth is negligible for an unlimited number of alternations.
- Region III, where size of the defect is rapidly increasing, is significant only for components subject to a few cycles of load (of the order of 10 or little more), as e.g. for the case of pressure vessels which are decompressed and compressed only a few times during their lives. Summarizing, parameters of fracture can be employed for the fatigue design, to estimate, for a component, the maximum defect size corresponding to the limit of its fatigue life.

9.7.1.1 Integration of the Paris Law and Its Criticality

The Paris Law can be integrate for intervals a in which the magnification factor M can be considered constant:

$$dN = \frac{da}{C \Delta K^m} = \frac{da}{C [M(a/W) \Delta \sigma \sqrt{\pi a}]^m}$$

$$dN = \frac{1}{C \cdot (M \Delta \sigma \sqrt{\pi})^m} \cdot \frac{da}{a^{m/2}}$$

integrating from an initial to a final value of a , the corresponding number of cycles is obtained:

$$N_f = \int_0^{N_f} dN = \frac{1}{C \cdot (M \Delta \sigma \sqrt{\pi})^m} \int_{a_0}^{a_f} a^{-m/2} da$$

then:

$$N_f = \frac{2}{C \cdot (m - 2)(M \Delta \sigma \sqrt{\pi})^m} \left(\frac{1}{a_0^{\frac{m-2}{2}}} - \frac{1}{a_f^{\frac{m-2}{2}}} \right) \tag{9.85}$$

Thus the formula offers the way to estimate the residual fatigue life of a component from an initial defect to the critical one, known the relative M factor.

Analyzing the formula Eq. 9.85, it becomes clear that the number of cycles to failure is an ill-posed result with respect to the variable a_0 . In fact, a small variation of this, leads to great variations on the number of cycles necessary to reach the critical dimension. Figure 9.38 illustrates this case for a numerical example taken as a reference.

It should therefore only be used for not too small values of the initial crack length. It is not possible e.g. to correctly describe the fatigue life for the values of the initial length corresponding to a newly formed crack or, as it was called, to a *visible* crack as result of a nucleation of low cycle fatigue.

It remains totally not treatable by Paris Equation the field of propagation of the small defects on which there is still a large specialized literature [4].

Other influence variables have not been taken into account and it could be necessary to investigate on them by the stress analysis point of view:

- Specimen configurations other than standards when they are chosen more geometrically similar to the cracked component.
- Residual stresses superimposed on the applied cyclic stress and interaction with other variables such as environment and heat treatment.
- Crack closure effect.
- Load sequence in case of variable amplitude loading.
- Small fatigue cracks.

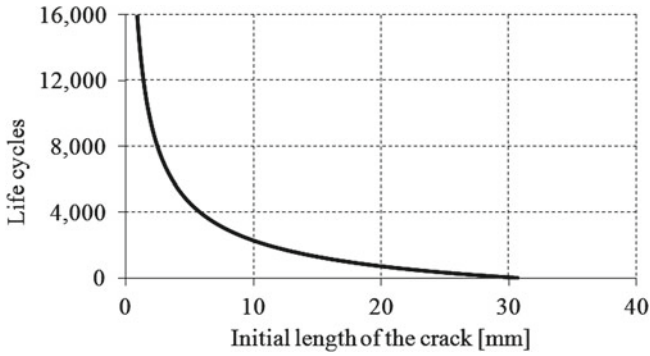


Fig. 9.38 Criticality of the Paris Law for small initial defects

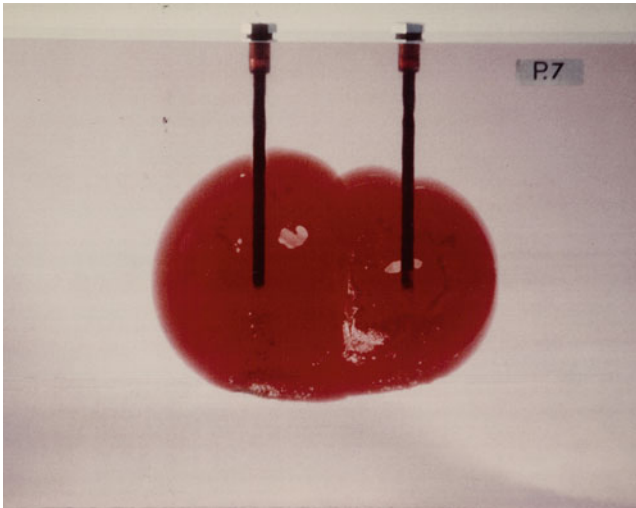


Fig. 9.39 Visual analysis of the propagation of two coplanar cracks in a plexiglass block [22]

9.8 Visual Analysis of Crack Propagation

The direct visual analysis in a transparent but elastic and isotropic material offers a valuable insight of the propagation phenomenon and a way to verify propagation laws. Plexiglass or epoxy resins are suitable for this purpose in the limit of their elastic behavior. In this case the internal pressurization is equivalent to uniform external state of stress, Fig. 9.10 (Figs. 9.39, 9.40 and 9.41). Fig. 9.42 shows the relationships between different units.

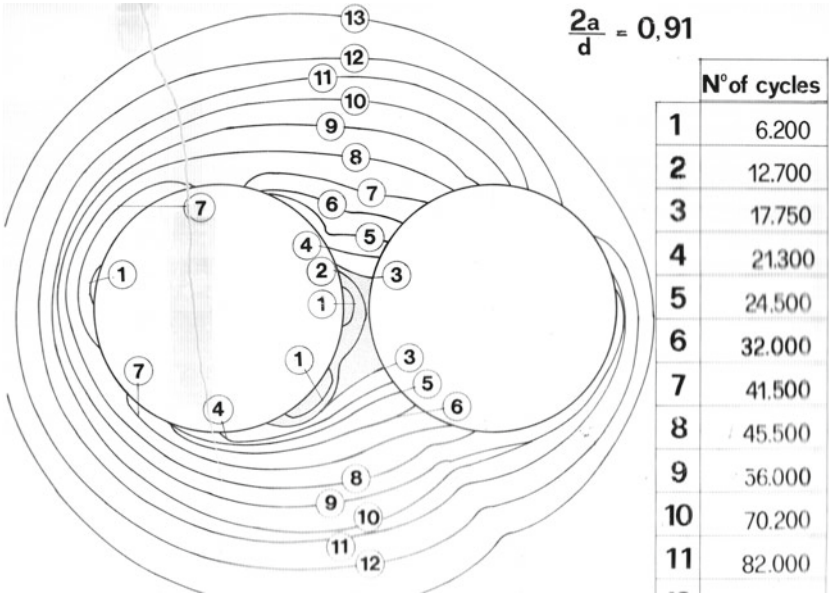


Fig. 9.40 Shape of the crack tips of two coplanar circular cracks in a plexiglass block

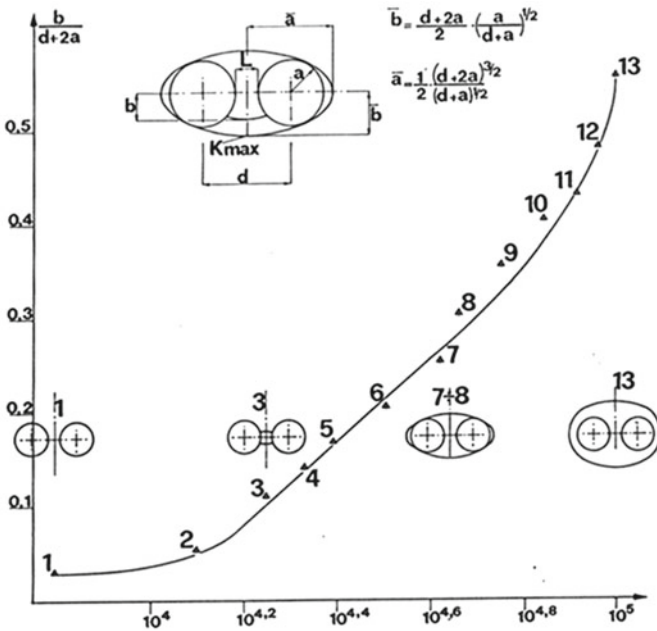


Fig. 9.41 Propagation crack tip front at different number of cycles

Variable	Formula	Symbol
Temperature	$T_C = (T_F - 32)/1.8$	FATT
	$T_K = (T_F + 459.67)/1.8$	FATT
ΔT	$\Delta T_C = \Delta T_K = \Delta T_F/1.8$	
Length	$1 \text{ inch} = 0.0254 \text{ m}$	B, W, a
Stress	$1 \text{ ksi} = 6,895 \text{ MPa}$	σ
Energy	$1 \text{ ft} \cdot \text{lb} \cdot \text{f} = 1.356 \text{ Joule}$	Charpy
	$1 \text{ Joule} = 1 \text{ N} \cdot \text{m}$	
	$1 \text{ ft} \cdot \text{lb} \cdot \text{f} = 0.173 \text{ kp} \cdot \text{m}$	
Toughness	$1 \text{ ksi} \sqrt{\text{inch}} = 1.1 \text{ MPa} \sqrt{\text{m}}$	K_{Ic}
	$1 \text{ ksi} \sqrt{\text{inch}} = 3.543 \text{ kp} \cdot \text{mm}^{-3/2}$	
	$1 \text{ MPa} \sqrt{\text{m}} = 3.22 \text{ kp} \cdot \text{mm}^{-3/2}$	
Force	$1 \text{ lb} = 4.448 \text{ N}$	F_Q

Fig. 9.42 Unit conversion

References

1. Barsom JM, Rolfe ST (1999) Fracture and fatigue control in structures: applications of fracture mechanics. ASTM manual series, vol 41. ASTM International, West Conshohocken. ISBN: 9780803120822
2. Eichlseder W (2005) Betriebsfestigkeit, vol 1. Montan Universitaet Leoben, Leoben
3. Ewalds HL, Wahnhill RJH (1986) Fracture mechanics. Arnold, Delft Uitgevers Maatschappij, London
4. Flewitt PEJ, Hayes PF, Schofield J, Smith RA, Turnbull A, Tomkins B, Wiesner CS (2002) Engineering structural integrity assessment: needs and provision. In: Proceedings 6th international conference on engineering structural integrity assessment, UK, 8–9 October, Manchester. EMAS Publishing, Manchester
5. Freddi A (1981) Fracture mechanics and testing of steel for large rotors. Int J Fatigue 3(2):71–76
6. Freddi A (1987) Integrity assessment of large forgings containing defects: a fracture mechanics approach. Int J Mater Prod Technol 2(3–4):328–337 (Interscience Enterpr Ltd, Geneve)
7. Inglis CE (1913) Stress in a plate due to the presence of cracks and sharp corners. In: Spring meetings of the 54th session of the institution of naval architects 14, vol 1, pp 1–3
8. Kobayashi AS (1993) Handbook on experimental mechanics. SEM. VCH Publishers, New York
9. Pook LP (1983) The role of crack growth in metal fatigue, vol 1. The Metal Society, London
10. Stephens RI, Fatemi A, Stephens RR, Fuchs HO (2002) Metal fatigue in engineering, vol 1, 2002nd edn. Wiley, New York
11. Timoshenko SP, Goodier JN (1969) Theory of elasticity. McGraw-Hill, New York
12. Dally JW, Riley WF (2005) Experimental stress analysis, 4th edn. McGraw-Hill Co, New York
13. Westergaard HM (1939) Bearing pressures and cracks. Trans Am Soc Mech Eng 61:A49–A53
14. Murakami Y (1987) Stress intensity factors handbook, vol 2. Pergamon Press, Oxford. ISBN 0-08-034809-2
15. Rooke DP, Cartwright DJ (1976) Compendium of stress intensity factors, vol 2, 1st edn. Procurement Executive, Ministry of Defence H.M.S.O, London (ISBN 0117713368: 0117713368)

16. Sih G (1973) Handbook of stress-intensity factors. Lehigh University, Institute of Fracture and Solid Mechanics, Bethlehem
17. Tada H, Paris PC, Irwin GR (2000) The stress analysis of cracks handbook, vol 1. ASME Press, New York
18. Chona R, Irwin GR, Sanford RJ (1985) Influence of specimen size and shape on the singularity dominated zone. In: Lewis JC, Sines G (eds) Fracture mechanics: fourteenth symposium, ASTM, STP 791, Theory and Analysis, vol 1, pp 3–23
19. Sanford RJ (1995) On the general solution of plane crack problems in Westergaard form. Computational mechanics. Springer, Berlin
20. Sharpe WN Jr, Sharpe WN (2008) Springer handbook of experimental solid mechanics. SEM, Springer (ISBN 978-0-387-26883-5)
21. Doyle JF (2004) Modern experimental stress analysis: completing the solution of partially specified problems, vol 1. Wiley, New York (ISBN 0-470-86156-8)
22. Freddi A, Camponuovo GF (1980) Analisi fotoelastica dello stato di tensione in prossimità di due difetti circolari complanari e in piani paralleli in una struttura caricata di forte spessore. VII Conv. Naz. AIAS, AIAS, vol 1
23. Diaz FA, Patterson EA, Siegmund P (2009) A new experimental method for calculating KI and KII using photoelasticity. In: (ed) Proceedings of the SEM 2009 annual conference, School Street Bethel, p 7
24. Wolfram S (2012) Mathematica Manual. Wolfram Research, 8 edn. <http://www.wolfram.com/mathematica/>
25. Curioni S, Freddi A, Dal Re V, Zannoni M (1986) Fracture mechanics analysis of a NiCrMoV steel for a turbine disk. Öiaz: Österreichische Ingenieur und Architekten Zeitschrift 1(8):320–324
26. Blumenhauer H, Pusch G (1982) Technische Bruchmechanik, vol 1, 1st edn. VEB Deutscher Verlag, Leipzig- Springer, Wien
27. Gdoutos EE (2010) Fracture mechanics, an introduction, vol 1, 2nd edn. Springer, Berlin
28. Curioni S, Dal Re V, Freddi A, Zannoni M (1986) Fracture mechanics analysis of a NiCrMoV steel for a turbine disc. ÖIAZ, Österreichische Ingenieuren und Architekten 1(8), Springer, Wien
29. Freddi A, Persiani F (1978) Holographic measurement of magnification factors for embedded interacting defects in transparent models. In: 6th international conference on experimental stress analysis, VDI, vol 1. VDI Berichte Nr. 313, UNI Muenchen (Germany)
30. Freddi A (1995) Fracture toughness measurement through a torsion test. ÖIAZ Zeitschrift des Österreichische Ingenieuren und Architekten 1(9):268–271
31. Freddi A, Curioni S, Caligiana G (1989) La vita a fatica torsionale negli alberi dei turbo gruppi. La collaborazione ENEA industria nel settore termo-elettromeccanico. other collaborators: Dal Re V, Cammarota GP, Cesari F, Maggiore A, Medri GL, Nicoletto G, Persiani F, Veschi D

Chapter 10

Stress Models in Biomechanics

Luca Cristofolini

Abstract In this chapter some applications of experimental stress analysis to biomechanical problems are presented, with special regard to the field of orthopaedics. First, an overview of the basic science applications is provided: this allows to understand the functioning of the musculoskeletal system. Models of the part or of the entire musculoskeletal system are built to describe, interpret and predict its function. Biomechanics also enables to measure the structural and mechanical properties of our musculoskeletal system in itself, and when incorporating an orthopaedic device. In this chapter a description is also provided of the experimental stress analysis tools applied in vitro and in silico to measure the most relevant mechanical quantities (forces, moments, strain, displacement, strength, mode of failure) in musculoskeletal structures. In the last part, a role of biomechanics closer to practical application is described, focusing on the design and validation of orthopaedic implantable devices. This issue is very relevant to manufacturers, practitioners and patients alike, and integrates a large number of biomechanical experiments as well as numerical simulations.

10.1 Introduction

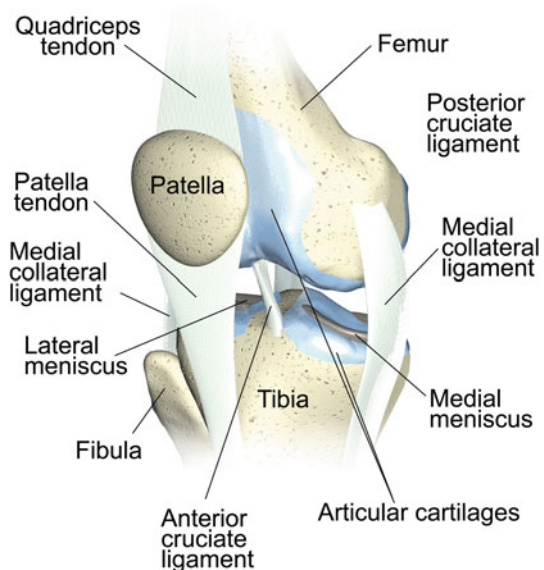
This chapter provides an overview of the possible applications of the methods of experimental stress analysis to biomechanical problems, focusing on orthopaedics. One possible classification of the applications of biomechanics relates to the scopes. Some activities are directed to the understanding of the basic principles of our musculoskeletal system (i.e. it relates to basic research). Other applications relate to the design and/or validation and/or optimization of biomedical implantable devices (i.e. it consists of applied research).

10.1.1 Biomechanics of the Musculoskeletal System

To understand the biomechanics of the musculoskeletal system, one must consider both its components, and how they interact with each other. While in the past such elements have in most cases been considered separately, the current state of our knowledge and the simulations recently developed make it possible (and almost mandatory) to take an integrative multidisciplinary approach, where each element is considered in its multiscale contexts.

First of all, let us analyze the structural components of the musculoskeletal systems, Fig. 10.1. The mechanical properties and the architecture of the tissues and organs involved are extremely important in determining the whole-body musculoskeletal biomechanics (movement, etc.). The muscles are the engines, with their ability of generating forces by converting metabolic energy in fibre mechanical contraction. The skeletal bones are the stiffest structures in our body. They serve as a scaffold and provide rigid leverage systems to convert muscle contraction into motion. The articular cartilages cover the extremities of the bones thus providing low-friction motion between interconnected bones. The ligaments link the bones to one another, avoiding excessive inter-segmental motion. The tendons link the muscles to the bones, and transfer the muscle (tensile) force to the bones.

Fig. 10.1 Example of structural elements of the musculoskeletal system: details of the knee



10.1.2 Methods for Measuring Musculoskeletal Kinematics and Dynamics

Models of the musculoskeletal system (and of its sub-systems) have been implemented with different strategies. This requires a combination with experimental measurements, to provide the input parameters, as well as to validate such models. Such experimental data are used both to be incorporated into numerical models, and as a diagnostic tool in itself (e.g. to evaluate the conditions of a patient). The measurement of the forces/moments is the background of any biomechanical model.

Measuring the musculoskeletal forces is an extremely difficult task. Direct measurement of the loads exerted by the muscles or transmitted by the tendons has been attempted for more than forty years in living subjects. Results have been quite frustrating, because of the difficulty of interposing an accurate transducer in the musculoskeletal system [1]. The first possibility that probably a mechanical engineer would consider consists in implanting a miniature transducer in the muscle or in the tendon. However, this is an extremely invasive solution, and offers poor accuracy because of the modifications caused by such implantation. Non-invasive tools such as ultrasonography, magnetic resonance imaging (MRI) are possible alternative options with less ethical and practical implications. However the latter approach is definitely less accurate.

The magnitude of the strain experienced by human and animal bones in real life during physiological activities has been measured by means of strain gages bonded to the bone itself [2, 3]. In other instances, instrumented staples and extensometers have been used to measure bone strain *in vivo* [4]. However, the actual loads transmitted by the bones are not directly measured this way. Moreover, these are very invasive techniques, involving dedicated surgery to expose the bone surface, and temporary attachment of a sensor onto the bone itself. For such reasons, this approach has been abandoned. Direct measurement of the internal loads becomes possible when a patient is treated with a prosthesis (e.g. hip or knee replacement). Telemetric prostheses have been designed and implanted, which enable measuring the magnitude and direction of the loads transmitted in an operated joint [2, 5–7]. This technique can obviously be applied only to patients undergoing joint replacement. Therefore, the measured loads are not necessarily representative of what happens in healthy subjects. Moreover, due to the high costs, data have been obtained on very small groups of patients. Despite such limitations, telemetrically measured joint loads [8–12] are the most extensive and reliable resource to understand how the musculoskeletal system works. Such information serves as an input to design pre-clinical validation methods for implantable devices, and to validate musculoskeletal models (see below). One of the most extensive databases of telemetrically measured joint loads at several anatomical sites has been made available by the research group of Bergmann and Rohlmann (<http://www.orthoload.com/>, Fig. 10.2). Alternatively, muscle activation can be measured in living subjects using electromyography

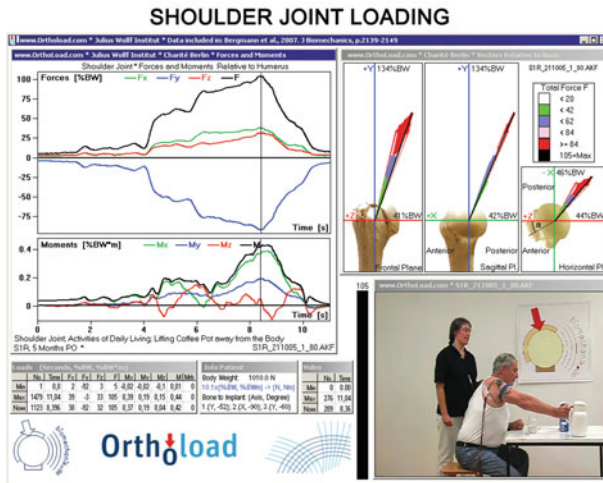
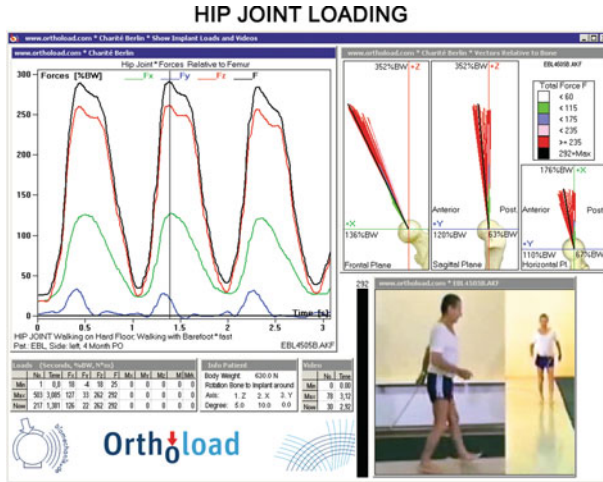


Fig. 10.2 Two examples of load profiles over time, which were measured in vitro using telemetric implants. In the top diagram the loading in the hip joint during gait (3 consecutive gait cycles) is reported. The bottom part shows the load components in the shoulder while lifting a cup (reproduced from <http://www.orthoload.com/>)

(EMG, [13]). EMG relies upon measurement of the electrical signal associated with muscle activity. As EMG measurement is unavoidably affected by large errors (e.g. noise, crass-talk between neighboring muscles), it is suitable to detect the timing of on-off activation of the muscles, rather than measuring the actual force exerted by a muscle.

Movement analysis by means of stereophotogrammetry allows gathering quantitative information about the kinematics of the musculoskeletal system during motor tasks, including an amount of directly measured quantities (e.g. motions and accelerations of the different segments, ground reaction forces), and numerical models describing the relationship between the segments. Marker locations are needed for the calculation of joint kinematics (e.g., inverse kinematics), which, in conjunction with the measured external forces (e.g. by means of multiaxial force platforms) are used for the calculation of joint loads, representing the resultant action of the muscles groups crossing that articulation, through inverse dynamics methods. EMG recordings are typically injected in this kind of models or used for validation purposes.

10.1.3 Biomechanical Musculoskeletal Models

Biomechanical models (and models in general) are simplified representations of a more complex reality. Because of their nature, one cannot take for granted that models predictions are representative of reality. For this reason, all models need a systematic, rigorous, quantitative verification and validation [14].

Accurate knowledge of the loads inside the skeleton during daily activity is extremely important for its scientific and clinical implications, as it may improve the diagnosis and treatment of various orthopaedics disorders. To analyze muscle and joint biomechanical function the following information is needed: external forces applied to the body, relative motion of the body segments, muscle and joint forces. Therefore, mechanical quantities (chiefly forces and motions) need to be measured as an input for computational models. The continuing improvement of computing power, together with more efficient and reliable modelling strategies, and accurate measurement of mechanical quantities (loads and motion) in living subjects is improving musculoskeletal models.

Biomechanical models driven by experimentally measured data enable investigating how the geometry of the muscles affects their effectiveness in generating moments around the articulations [15, 16], how the lengths and lever arms of skeletal muscles are altered by surgical procedures [17], or the ability of muscles to accelerate the body segments during movement [18]. In all such cases, the most common and challenging part relates to the estimation of loads transmitted by the muscles and joints during movement [19, 20]. Possible impacts of such biomechanical models include a better understanding of the events leading to bone fracture [21], and the prediction of clinical outcomes after implantation of a joint prosthesis (to estimate joint function, implant stability, bone adaptation to an altered strain distribution) [22, 23].

What should be included in a multibody dynamic model of the musculoskeletal system depends on the scopes of the model itself. If the goal is to analyze the kinematics and the skeletal loads, bone segments can be described as infinitely rigid

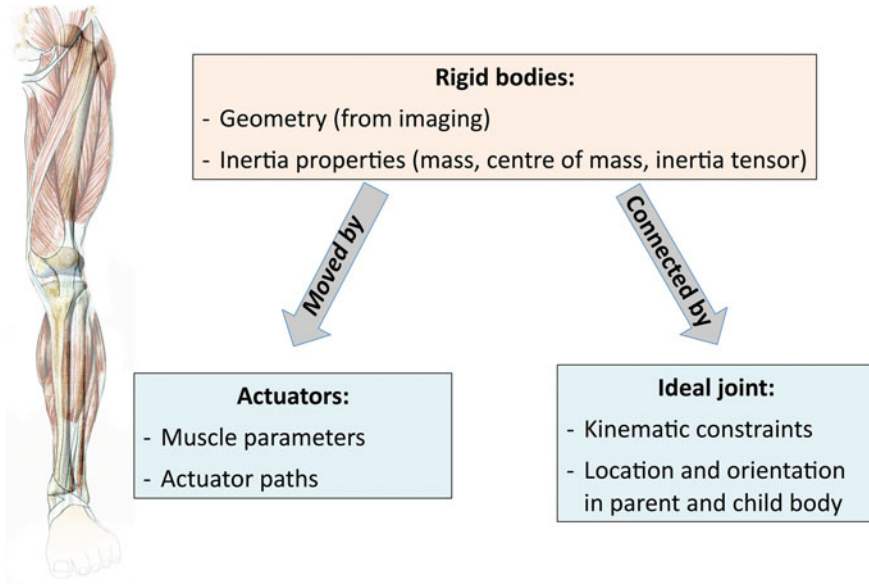


Fig. 10.3 Main parts of a musculoskeletal model for simulations of the dynamics of human movements. Courtesy of Giordano Valente (Rizzoli Orthopaedic Institute, Bologna, Italy)

bodies (bone tissue is orders of magnitude stiffer than any other living tissue), moved by the muscles, and connected by ideal joints, Fig. 10.3. Measurement of the input parameters (i.e. model identification) requires several complex operations on the specific subject, unless generic models based on average anatomy from population studies are adopted. The most common paradigm for the prediction of muscle and joint forces in human movement adopts a combination of musculoskeletal models, ground reaction forces (measured by multi-axial force platforms), and forward or inverse dynamics. In forward dynamics methods, muscle forces are estimated by integrating the equations of motion forward in time, using muscle excitation signals as a driver. This approach has high computational costs without significant advantages over the inverse dynamics approach. A more efficient implementation of the forward dynamics method is to follow a parameter optimization strategy: muscle loads are estimated to minimize a cost function, while targeting an experimentally measured kinematics dataset.

Inverse dynamic models rely on gait stereophotogrammetric measurements, applied to algorithms to calculate the articular moments associated with such kinematics. Muscle loads are then estimated by applying optimization methods to solve an indeterminate problem (agonist and antagonist muscles make the problem over-determined), under the hypothesis of some optimal performance criterion (e.g. uniform stress, or minimum metabolic cost).

10.2 Biomechanical Investigations in Orthopaedics at Different Dimensional Scales

Implantable devices have originally been tested pre-clinically *in vitro*, long before computer models could provide reliable and relevant information. The main strength of *in vitro* testing is that a physical specimen (for instance incorporating a bone segment and a prototype of the implantable device) is tested under application of real loads. Therefore, *in vitro* experiments closer to reality than numerical models.

Mandatory tests (often described by international standards) focus on the implantable device alone. More complex *in vitro* experiments focus on the biomechanical function of the device when implanted in the host bone: in these cases the test needs to include one or more bone segments. Bone segments for such tests can either be human cadaveric specimens, or animal bones, or artificial bone replicas. Artificial bone models (usually made of composite material to mimic the specific properties of bone tissue) have the advantage of being available in large numbers, easy to handle, and highly reproducible. In addition, bonding a transducer (e.g. a strain gage) to a synthetic bone is much easier than with a real bone specimen (which incorporates organic fat molecules and needs to be kept hydrated). The strengths of synthetic models can also be seen as a limitation: their high reproducibility makes them unsuitable to represent the variability among subjects. Animal tissue specimens are relatively easy to obtain through the food supply, and have mechanical properties that are to some extent comparable to human bone. In most cases the anatomy and the structure of animal specimens are different from human bone because of the different biomechanics of humans (with bipedal locomotion) compared to such animals (quadrupeds in most cases). For these reasons, artificial bone models and animal specimens are suitable for the first stages of validation, but are unsuitable for more refined studies where similarity to the human is fundamental. Cadaveric human tissue specimens are difficult to obtain. Finally, correct preservation of specimens of human or animal is extremely critical [24, 25].

Biomechanical testing of bone segments involves application of relevant loads to the specimen, while applying well-defined constraints. A critical issue is proper identification of the load magnitude and direction to be applied, as the loading conditions can vary significantly, Fig. 10.4. In some cases a simplified loading scenario is to be preferred: load components are applied singularly (e.g. a pure compressive force, or a pure torsional moment, or a pure bending moment). In other cases more complex loading scenarios are simulated either to replicate physiological motor tasks or a trauma loading, by incorporating more load components [26].

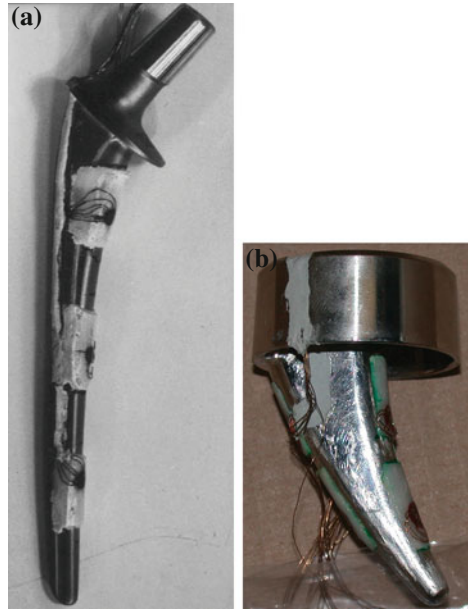
The *in vitro* test includes a number of transducers for measuring the relevant biomechanical quantities. First of all, load cells measure the applied loads. Strain on the bone surface can be measured means of strain gages, Fig. 10.4. Due to the complex geometry, and the anisotropic nature of the tissues under investigations, the principal directions of strain are unknown. Therefore, the use of triaxial strain gages



Fig. 10.4 Examples of in vitro biomechanical experiments to measure the stiffness, strain distribution and strength of intact and operated bone segments: **a** four-point-bending of a human tibia; **b** single-leg-stance loading of a femur implanted with a hip stem; **c** sideways fall simulation on a human femur; **d** compressive testing of an isolated vertebra and **e** compression-bending of a three-vertebrae spinal segment. The strain gages that were used for measuring strain on the bone surface are visible (Copyright of the VPHOP consortium; reproduced with permission)

is mandatory to enable meaningful measurement of the principal components of strain and their direction. In addition, to measure internal strain, strain gages may be embedded inside the layer of acrylic cement surrounding an implant, Fig. 10.5 [27]. Alternatively, the strain distribution can be measured using digital image correlation (DIC) [28]. DIC allows full-field measurement of displacements and of the strain distribution, with larger noise and error than strain gages. In addition, to elucidate the failure mechanism and the point of fracture initiation, high-speed videos (with frame rates of the order of 10,000–20,000 frames per second, Fig. 10.6) can be extremely useful [29]. Experimental measurements in biomechanical applications are affected by a number of limitations that in some cases make them difficult to implement or insufficient. The first, obvious limitation is that experimental measurements are affected by both random and systematic error: such errors may become large when biomechanical structures are tested, which have an irregular geometry and complex material properties. In addition, biomechanical testing often requires complex setups and several measurements: obtaining any additional single information from an in vitro experiment typically involves adding more transducers, more actuators, more

Fig. 10.5 Strain gages prepared on a cement layer surrounding a standard hip stem (a) and a resurfacing hip prosthesis (b): the stem is then inserted thus embedding the strain gage inside the cement layer, to enable measurement of internal strains



controllers and more data loggers, with the associated costs. Moreover, there is little scale economy if several samples need to be compared such as when different loading scenarios need to be explored, or when different versions of an implantable device are tested. Because of such limitations, numerical models are often seen as an alternative to *in vitro* biomechanical experiments.

The considerations above to some extent may push towards numerical modelling. In most cases Finite Element (FE) models are used to predict the biomechanical response (state of stress/strain, stiffness, strength, etc.) of orthopaedic constructs. However, the intrinsic limitations of numerical models still apply, also in a biomechanical context. First of all, numerical models (like indeed any sort of model) are unverified hypotheses (that is: opinions). From an epistemological point of view, a model can never be proven true, but can only be falsified [30]. In the best case, a numerical model can be validated by checking that the predictions have a sufficient accuracy [14]. Validation must be performed by comparing the numerical model against experimentally measured magnitudes under specific loading conditions. Challenging the model under the most critical and diverse conditions involves applying several loading scenarios to a physical specimen, and to the corresponding numerical model. Mechanical magnitudes (forces, strain, etc.) must be measured and qualitatively compared to quantify the fidelity of the numerical model to the validation experiment. A validated numerical model is by no means a shortcut: it is expensive, it requires a variety of facilities and know-how, and it involves significant effort before reasonable agreement is obtained. Once available, a validated numerical model is an extremely powerful tool.

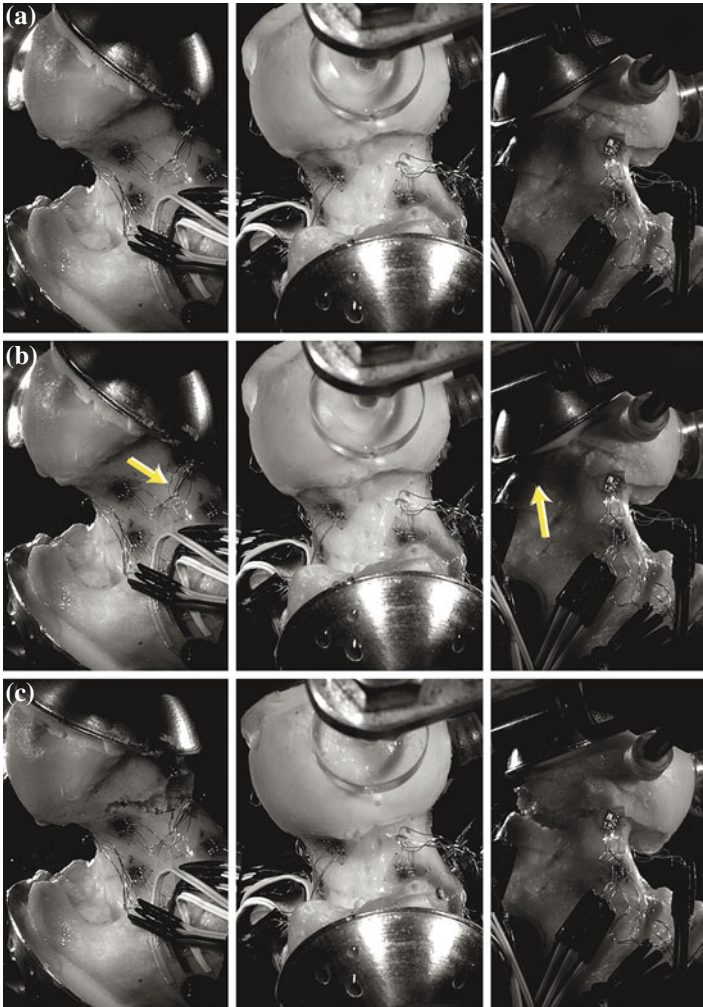


Fig. 10.6 Fracture mechanism observed during a simulated sideways fall observed with in the high-speed videos (12,500 frames/s). The image in the centre of each picture is a direct view of the femoral neck from the lateral side; the ones on the *left* and *right* are reflected images (posterior and anterior sides respectively) obtained from the two mirrors placed next to the femur and suitably oriented. Picture **a** shows the femur shortly before the first signs of fracture are seen (0.3 ms before Picture **b**). Picture **b** shows the instant when compression failure is seen on the superior-lateral side (indicated by the *yellow* pointers). Picture **c** (70 ms after Picture **b**) shows the final stage, when the medial sides fails (in tension). The pictures have low resolution (1 pixel = approximately 0.2 mm on the physical specimen) because they were acquired by the high-speed camera (Copyright of the VPHOP consortium; reproduced with permission)

As discussed above, both biomechanical experiments and numerical models suffer from intrinsic limitations: verifiability, accuracy and fields of application [31]. Direct in vivo measurements performed on living subjects are in principle the most desirable option in scientific terms, as they are close to the physical reality. However, because of practical and ethical limitations discussed above, data available from in vivo measurements is limited. In vitro experiments enable more extensive simulations, including, more extensive and better accurate measurements, and even destructive testing. Experiments (whether in vivo or in vitro) can support and improve numerical models along three paths, Fig. 10.7. First of all, in vitro measurements enable a preliminary identification of the relevant failure scenario(s) to be modelled. In a second instance, in vitro measurements enable the identification of the model parameters (e.g. material properties, friction coefficient, etc.). Finally, in vitro experiments support a quantita-

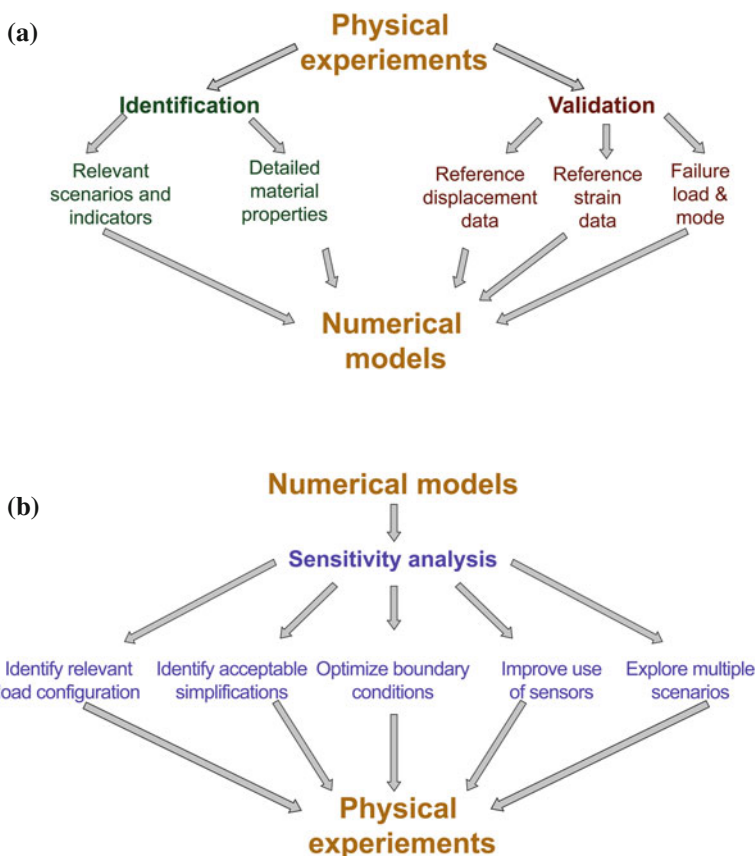


Fig. 10.7 Pathways along which numerical simulations and in vitro experiments should complement each other: **a** flow of information from in vitro experiments to numerical models; **b** role of numerical models in improving in vitro experiments

tive validation of the model based on mechanical magnitudes (strain, displacements, micromotions, fracture, etc.) directly measured on a physical specimen [32].

At the same time, numerical models can be an invaluable resource to improve in vitro experiments, Fig. 10.7. First of all, numerical models can be used to identify the most relevant loading configurations among a number of motor tasks that cannot be practically/economically replicated in vitro. Numerical models can then help identify which simplifications are acceptable for the in vitro simulations. Sensitivity analyses supported by numerical models can also help optimize the use of transducers to provide measurements at the most relevant locations while minimizing errors. Finally, numerical models are more effective than in vitro experiments in order to explore a variety of different conditions (e.g. material properties, interface conditions [33] etc.)

Therefore, an integration of numerical models and in vitro experiments is desirable as it provides more reliable and more exhaustive information than any of the two approaches taken individually. To exploit the synergies at best, experiments and numerical simulations should be designed together [31]. To prevent possible bias, experiments and numerical simulations should be carried out independently, in a double-blinded fashion [34]. Comparisons must be based on quantitative indicators, with clearly defined metrics and thresholds.

10.3 Case Study: Strain Distribution in the Human Tibia, a Uniform Stress Structure

Most biomechanists and evolutionists agree that bones are structured to resist daily loads in an optimal way. One of the main loads components acting in the tibia during daily motor tasks is a cantilever load. In fact, the antero-posterior load component acting during locomotion causes a linearly varying bending moment. We investigated if the cross-section of the shaft of the tibia, and its variation along the tibia make it an optimized structure with respect to such loads [35]. The geometry and material properties were extracted from the computed tomography, and analyzed. A linear variation along the tibia was found for the second moments of area and inertia, and for the section modulus. This suggests that the structure is optimized to resist a bending moment that varies linearly along the tibia. Each of the six tibias was prepared with 28 triaxial strain gages (84 strain grids on each tibia) equally spaced along the tibia, on its anterior, posterior, medial and lateral sides. The tibias were loaded in the elastic range under cantilever loading in the sagittal and frontal planes, under quasi-constant-bending in the sagittal and frontal planes, under torsional loading, and with an axial force. The strain distribution was remarkably uniform when cantilever loading was applied in the sagittal plane, Fig. 10.8 and slightly less uniform when cantilever loading was applied in the frontal plane. Strain variations were one order

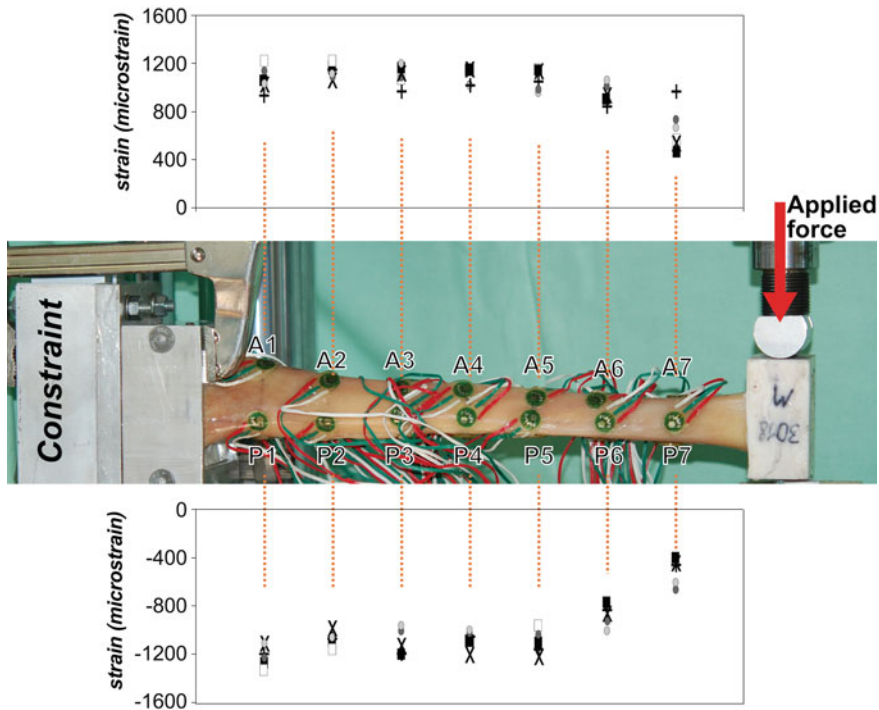


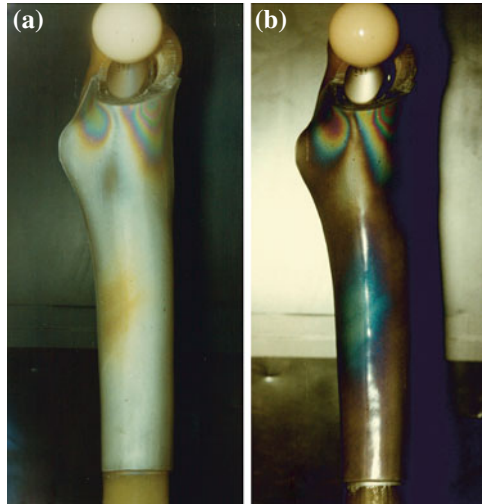
Fig. 10.8 Tibias were loaded in a cantilever fashion (the proximal part, on the *left*, was constrained while a load was applied perpendicular to the tibia at the distal end). The strain plots for six specimens show a rather uniform strain distribution

of magnitude larger for all other loading configurations. Therefore, these in vitro strain measurements confirm that the tibia acts as a uniform-stress structure when subjected to daily loads which generate a cantilever force.

10.4 Case Study: Bone Strains Caused by Press Fitting of a Hip Stem

Some hip stems are inserted in the host femur with no means of fixation other than friction due to press fitting. In some cases this causes intraoperative bone fractures. The aim of this study was to investigate the magnitude and distribution of the strain/stress caused by press fitting [36, 37]. Composite femurs were prepared by an experienced surgeon to host a cementless hip stem. The stem was firmly press fitted in place. A photoelastic coating (2 mm of araldite) were applied with a reflective glue. When the stem was extracted, the state of stress induced by press fitting was removed and

Fig. 10.9 This composite femur was prepared with a photoelastic coating to measure the strains induced by insertion of a press-fit hip stem. The full-order (a) and half-order (b) isochromatic fringes highlight the strain distribution induced by press fitting



isochromatic fringes appeared, Fig. 10.9. The state of strain due to removal of the stem is identical in magnitude (and of opposite sign) to that caused by press fitting.

10.5 Case Study: Deformations of an Acetabular Cup by Means of Holographic Interferometry

Most orthopaedic devices consist of an assembly of two or more components. Deformations of implantable component sometimes play a crucial role for their survival. As an example, the deformations of the metal backing of the acetabular cup of a total hip replacement were investigated, as they can play a role in the failure mechanism of the insert, both due to contact stress concentration, and fretting. Holographic interferometry was chosen as it allowed: (i) investigating the full field of displacements and strains in 3D, and (ii) was capable of measuring the extremely small displacements expected in this case, Fig. 10.10. Images were then digitally processed to compute the field of displacements and strains [38].

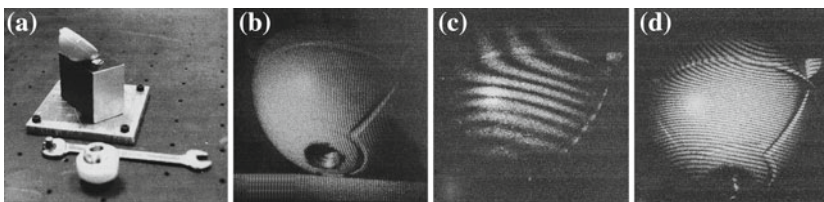


Fig. 10.10 Measurement of the displacements in an acetabular cup by means of holographic interferometry: a specimen; b rigid-body rotation (0.015°); c 33% load; d 100% load

10.6 Biomechanical Risk Assessment: Experimental Stress Analysis and Modelling

To compensate, either in a temporary or permanent way, to different diseases affecting the musculoskeletal system a number of diverse devices have been designed in the past hundred years, Fig. 10.11. Because of the many biomechanical challenges, the focus in this section will be on implantable devices. Total joint prostheses will be taken as a representative instance. Currently, joint prostheses are implanted in an enormous number of patients: in the 30 member countries of the Organisation for Economic Co-operation and Development (OECD) the rate varies between 0.50 and 1.40 hip replacements per year per 1000 inhabitants [39], corresponding to a total of well over one million hip replacements per year. Similar, and increasing trends are observed for other joint prostheses such as the knee, shoulder, ankle and elbow. Extreme caution is needed before a new prosthetic design is introduced to the market: the risk associated with known failure scenarios should be equal to, or possibly lower than that observed in devices already in use, while providing some additional benefit/feature. The validation paradigm includes extensive numerical and/or experimental tests (pre-clinical validation), a clinical trial on a controlled small and

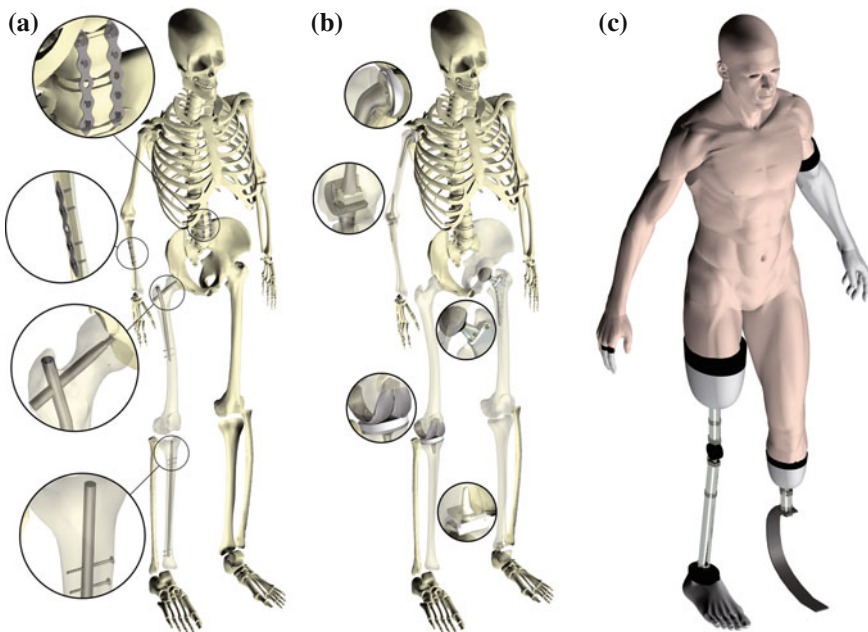


Fig. 10.11 Examples of orthopaedic devices. **a** Fracture fixation devices: spinal fixation, radial plate and screws, femoral screw and plate, tibial intramedullary nail. **b** Articular prostheses: total shoulder, elbow, hip, knee, and ankle replacements. **c** External prostheses: arm, finger, and leg above and below the knee

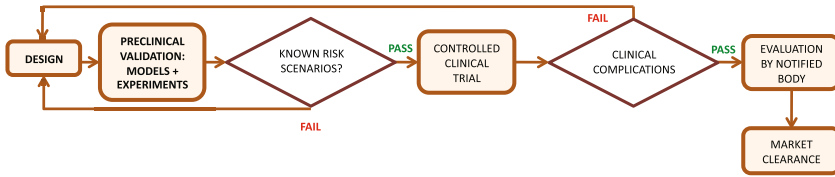


Fig. 10.12 Schematic of the path from when an implantable device is first conceived, to introduction to widespread clinical use

selected set of patients (clinical trial) and eventually commercialization of the device, Fig. 10.12. The valuable load data recorded during physiological motor tasks using telemetric prostheses [8–12] provide fundamental information on the biomechanical conditions in which joint replacements must safely withstand operate. A balanced mix of experimental studies on synthetic bone analogues [40, 41], animal models [26, 42] and human cadaver preparations [43, 44] enables thorough investigation of the most relevant failure scenarios. At the same time, biomechanical computational models allow an integrated approach where fully validated models are combined with experiments so as to explore every possible failure scenario [45, 46].

There are two fundamental issues that have emerged over the years that need to be addressed, which may jeopardize further innovation in joint replacement. First: considering the extremely good outcome of some joint replacement (the best hip joint prostheses fail in less than 2–3% of cases over 10–15 years), is it possible to improve? Secondly: considering all the known failure scenarios, and assuming that the risk of each of them should be assessed during the pre-clinical validation stage, the validation cost would become unbearable. Can we decide what risks need to be addressed and which ones can be neglected? To achieve the best possible validation for a given budget/time frame, the state of the art process combines formal analysis, numerical modelling, and in vitro experiments.

10.6.1 Risk Analysis

In order to obtain approval for widespread use of a medical device, the manufacturer must prove according to a well-established paradigm, the lack of risk and possibly the improvement compared to existing devices. For instance, to obtain CE approval for use within the European Union, the device must be analyzed according to the ISO 14971:2007 Standard (Application of risk management to medical devices). Similarly, to introduce a medical device to the USA market, clearance must be obtained from the FDA by demonstrating that all foreseeable risks are avoided, and the device is similar or better than existing ones. There are various methods to perform a risk analysis; all follow a common scheme [47]. First of all, a multidisciplinary team that should include regulatory experts, engineers, and clinicians must identify potential hazards. Then, the probability of occurrence and the risk must be identified for each hazard, using either a top-down Fault Tree Analysis (FTA), or a bottom-up Failure Mode and Effect Analysis (FMEA). Then, the most critical hazards must be identified

and mitigated until the risk level is low enough to be acceptable. The risk analysis starts from considering the known clinical failure scenarios (the undesired effects in the FTA approach). These are usually identified from previous experience, ruling out scenarios that are definitely not relevant to the device under investigation. For instance femoral neck fracture is a known complication of epiphyseal replacements, but it is not relevant for conventional hip prostheses where the femoral neck is removed. Each clinical failure scenario must then be associated with the corresponding mode(s) of failure that could generate such a failure scenario. The FMEA approach highlights only the primary associations. In principle, top-down approaches should use the more complex FTA approach, which can establish all the combinations of failure modes that can produce an adverse effect (for instance, a negative synergy of insufficient primary stability and wear can induce implant loosening). This is clearly a more rigorous approach; unfortunately, in most cases it cannot be implemented because of the lack of basic information. Even if a simple FMEA approach is taken, the identification of the clinical failure scenarios that are associated with the failure modes are the basis of the prevention of failures. The failure modes should then be used to design the set of tests/simulations for the pre-clinical validation. As the validation is limited in time and cost, it is ethical to use such resources to address the failure modes that according to the risk analysis are most critical. It is important to note that, no matter how carefully and extensively the risk analysis is carried out, it cannot predict the unpredictable: if a device is totally innovative, it could induce a type of failure scenario that has never been observed before.

Three elements of the bone with an implanted prosthesis could fail under cyclic load: the prosthesis, the fixation material (typically acrylic bone cement, if present), and the host bone itself. Thanks to the severe mandatory tests, mechanical failure of the device in itself is an extremely rare event (yet not impossible [48]). Fretting corrosion may occur in modular solutions, especially when two different metals are coupled: fretting analyses are recommended [49], under worst case scenario chemical and mechanical conditions [50]. Acrylic bone cement (typically polymethylmethacrylate, PMMA) poses completely different issues: the cement itself is frequently the weak link of the chain. Thus, it is important to determine the state of stress/strain inside the bone cement. Strain gages embedded inside the cement can provide reliable strain measurements at selected locations, Fig. 10.5 [27]. Therefore, it can be advantageous to use this experimental information to validate a model, and use the model to predict the entire stress field in the cement. Estimating the risk of bone fracture by numerical models is very difficult as the ultimate properties of bone vary greatly. This uncertainty is due to a number of factors, including the inability of actual modeling techniques to recognize small features on the bone surface where the initial crack may occur and lead to failure [51]. Thus, an experimental validation is preferable in this case. In addition, the presence of the implant (which is orders of magnitude stiffer than the host bone) generally causes an alteration of the stress distribution known as stress shielding. This can cause adverse adaptation of the bone tissue (loss of bone stock). In order to exclude severe bone resorption over time, the strain distribution in the implanted bone should not be too different from that in the intact one [52, 53].

10.6.2 Case Study: The Paradigm for Pre-clinical Validation of New Implantable Devices

In these pages the path to thoroughly validate a new implantable device is illustrated using a specific paradigmatic example: a total hip prosthesis. In most cases, new devices can be compared to the similar devices already on the market (the so-called predicate devices). So, the focus on the biomechanical validation should be on the aspects on which such a device may be different (advantageous and/or more risky) than known ones. For instance, the claim of a new hip prosthesis could be a reduced risk of fracture of the implanted bone. The risk analysis must be performed. The most commonly reported clinical failure scenarios for this type of device are identified, and various failure modes are associated to each of them, Fig. 10.13. The validation should combine the best of in vitro tests and numerical models. Validation must use a set of human cadaveric bones as a test bench, both for the mechanical tests, and for the numerical models. The bones must be scanned with computer tomography (CT) [54]; the CT dataset is used to create a finite element model [55]. The physical specimens must then be instrumented for measurement of the relevant mechanical magnitudes (e.g.: strain, displacements, relative micromotions, fracture), and subjected to multiple load cases. The in vitro mechanical measurements will complement

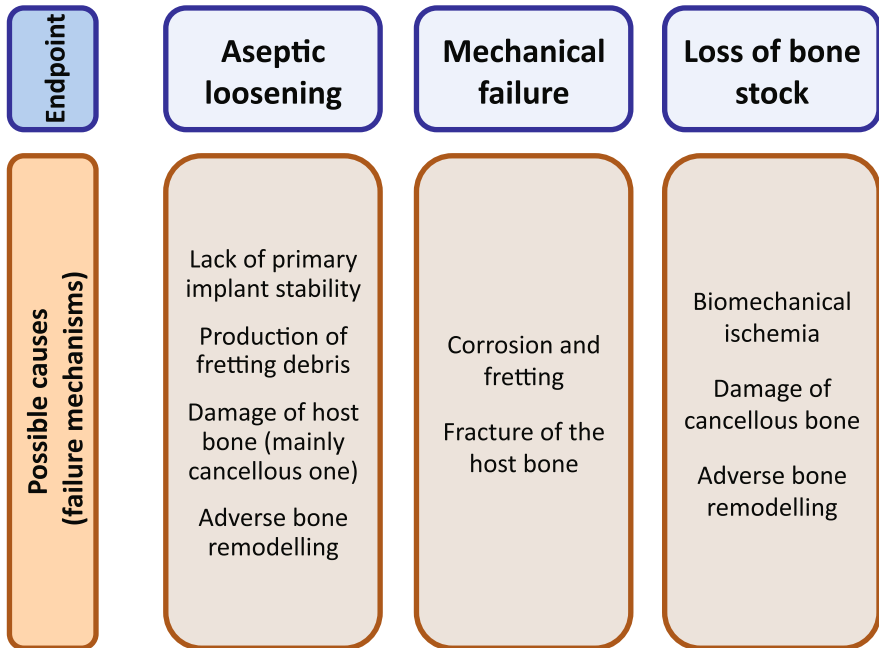


Fig. 10.13 Overview of common clinical failure scenarios, together with the various failure modes: three cases are reported, which are typical for a femoral resurfacing prostheses

the predictions obtained by the model under the same load cases, so as to provide a validation of the FE model.

For each failure mode a numerical model or an experimental test is designed to assess the risk of occurrence of such a failure, for this specific new device. To investigate the risk of fracture, the intact bone and the implanted bone are subjected to multiple load cases, mimicking the peak loads observed during relevant activities, while the strain distribution is measured in the host bone. The focus must be on the load case, which induces the highest risk of fracture in the implanted bone, which can be assessed in terms of strain distribution (in fact, bone fracture is driven by a strain-based failure criterion [56, 57]), Fig. 10.14. A surgeon applies the surgical procedure that is considered standard for that device to deliver the prosthesis to the cadaver bones. In many instances it is preferable to use paired contralateral

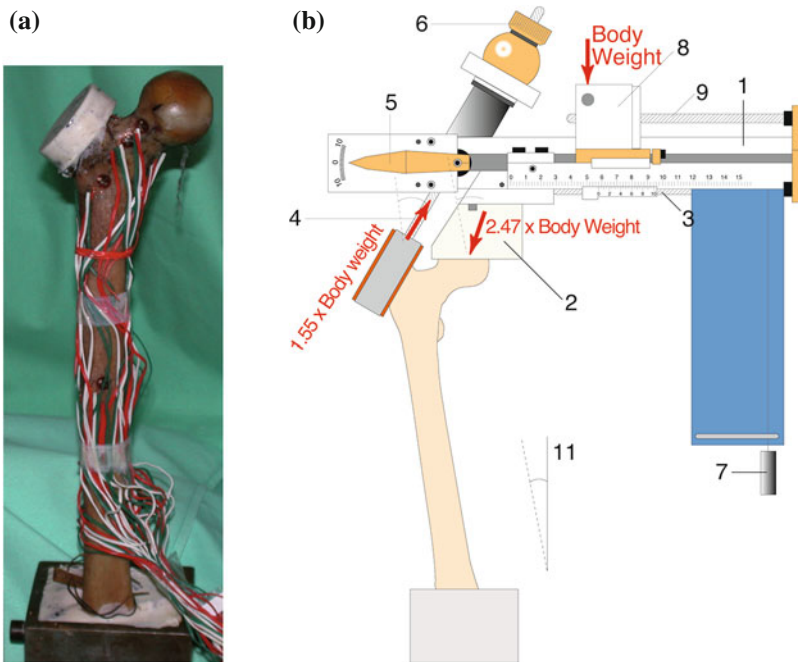


Fig. 10.14 Mechanical test to measure strains on the bone surface by means of strain gages to detect deviations between the intact femur, and the same femur after implantation with a hip prosthesis. **a** Anterior view of a femur prepared with 16 triaxial strain gages, a cylindrical block on the greater trochanter (for application of the gluteal load), and of the distal pot to hold them aligned during the tests. **b** Loading set-up to replicate the loads during a one-leg stance. The direction and magnitude of the hip and abducting forces are such as to balance the resultant force applied to the top of the cantilever. Their ratio and direction is controlled by adjusting the lever arms. The following components are connected to the cantilever (1), a movable hip socket (2), whose position can be adjusted with the screw (3), an abducting rod (4), free to pivot together with the angle indicator (5), its length is adjusted by tightening the bolt (6), a plumb line (7), indicates if the cantilever (1) is horizontal; a slide (8), is adjusted by means of the screw (9)

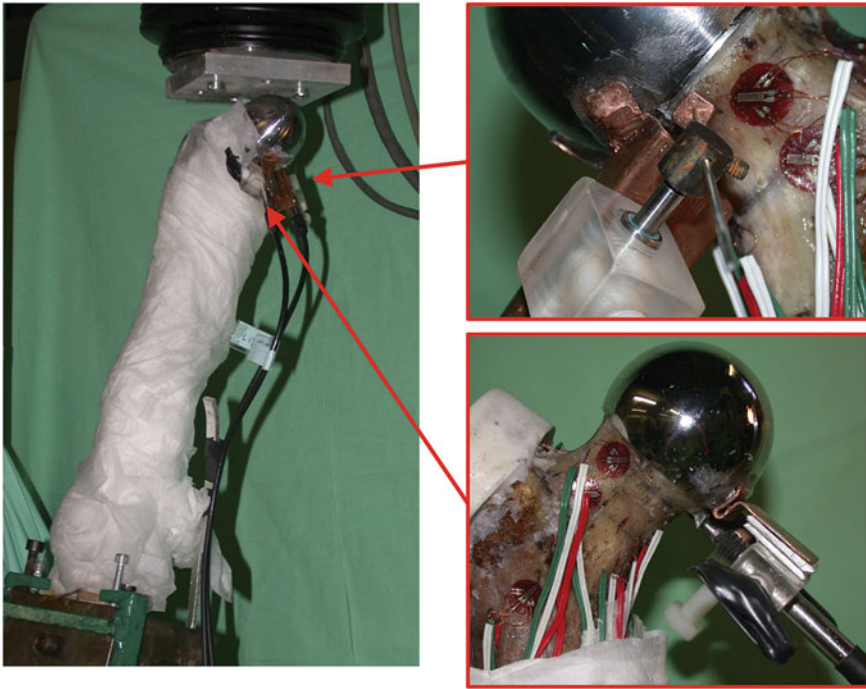


Fig. 10.15 Example of loading setup to measure implant-bone micromotions for a resurfacing hip prosthesis by means of high precision LVDTs (overview and details)

bones to enable comparison against a control (either intact, or implanted with a well-established similar device to provide a baseline value). The mechanical stability of the implanted prosthesis in the host bone can be investigated with an FE model simulating implant-bone contact, or, more reliably, with a mechanical test. Bone-implant and cement-implant inducible micro-movements are typically measured by means of LVDTs under different loading conditions, Fig. 10.15. Normally cyclic loads are applied so as to measure the inducible micromotion (elastic movement at every cycle) and permanent migration (settling that accumulates cycle after cycle). The typical precision of such relative micromotion measurements can be as high as few microns, which is sufficient to predict the risk of implant loosening [58, 59]. When there is concern about the integrity of the cement mantle surrounding the implants, the implant can be inspected by means of dye penetrants to highlight the presence of cracks after cyclic loading Fig. 10.16.

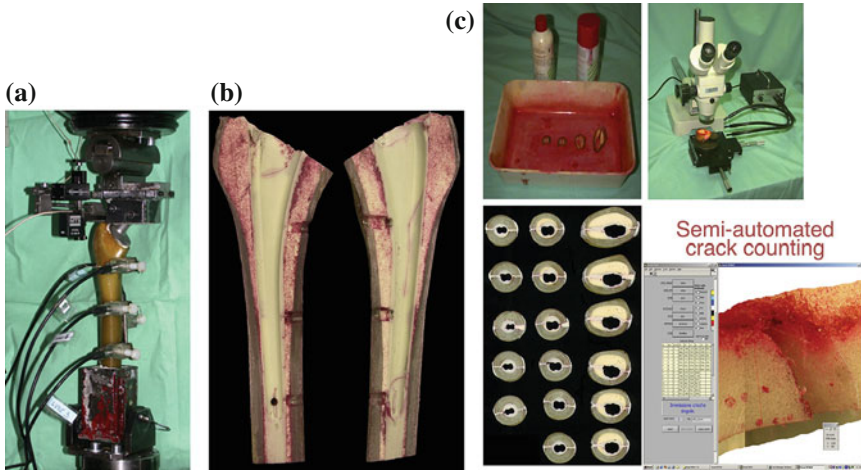


Fig. 10.16 Example of cement mantle surrounding a hip prosthesis sectioned after mechanical cyclic testing in vitro (a). Dye penetrants are used to highlight the presence of cracks in the whole mantle (b), and in its cross-sections (c)

References

1. Fleming BC, Beynon BD (2004) In vivo measurement of ligament/tenon strains and forces: a review. *Ann Biomed Eng* 32(3):318–328
2. Caler WE, Carter DR, Vasu R, McCarthy JC, Harris WH (1982) In vivo intracortical loading histories calculated from bone strain telemetry. In: Huijskes R, van Campen DH, de Wijn JR (eds) *Biomechanics—principles and applications*. Martinus Nijhoff Publishers, Boston
3. Lanyon LE, Hampson WGJ, Goodship AE, Shah JS (1975) Bone deformation recorded in vivo from strain gauges attached to the human tibial shaft. *Acta Orthop Scand* 46:256–268
4. Yang PF, Bruggemann GP, Rittweger J (2011) What do we currently know from in vivo bone strain measurements in humans? *J Musculoskelet Neuronal Interact* 11(1):8–20
5. Davy DT, Kotzar GM, Brown RH, Heiple KG, Goldberg VM, Heiple KGJ, Berilla J, Burstein A (1988) Telemetric force measurements across the hip after total arthroplasty. *J Bone Jt Surg Am* 70:45–50
6. Graichen F, Bergmann G (1991) Four-channel telemetry system for in vivo measurement of hip joint forces. *J Biomed Eng* 13:370–374
7. Taylor SJG, Perry JS, Meswania JM, Donaldson N, Walker PS, Cannon SR (1997) Telemetry of forces from proximal femoral replacements and relevance to fixation. *J Biomech* 30:225–234
8. Bergmann G, Deuretzbacher G, Heller M, Graichen F, Rohlmann A, Strauss J, Duda GN (2001) Hip contact forces and gait patterns from routine activities. *J Biomech* 34(7):859–871
9. Bergmann G, Graichen F, Rohlmann A (2004) Hip joint contact forces during stumbling. *Langenbecks Arch Surg* 389(1):53–59
10. Kirking B, Krevolin J, Townsend C, Colwell CW, D’Lima DD (2006) A multiaxial force-sensing implantable tibial prosthesis. *J Biomech* 39(9):1744–1751
11. Taylor SJ, Walker PS (2001) Forces and moments telemetered from two distal femoral replacements during various activities. *J Biomech* 34(7):839–848
12. Taylor SJ, Walker PS, Perry JS, Cannon SR, Woledge R (1998) The forces in the distal femur and the knee during walking and other activities measured by telemetry. *J Arthroplast* 13(4):428–437

13. Bronzino JD (2006) *The biomedical engineering handbook*, 3rd edn. CRC: Taylor & Francis, Boca Raton
14. Babuska I, Oden JT (2004) Verification and validation in computational engineering and science: basic concepts. *Comput Methods Appl Mech Eng* 193(36–38):4057–4066
15. Lenaerts G, De Groote F, Demeulenaere B, Mulier M, Van der Perre G, Spaepen A, Jonkers I (2008) Subject-specific hip geometry affects predicted hip joint contact forces during gait. *J Biomech* 41(6):1243–1252
16. Valente G, Martelli S, Taddei F, Farinella G, Viceconti M (2012) Muscle discretization affects the loading transferred to bones in lower-limb musculoskeletal models. *Proc Inst Mech Eng H* 226(2):161–169
17. Delp SL, Ringwelski DA, Carroll NC (1994) Transfer of the rectus femoris: effects of transfer site on moment arms about the knee and hip. *J Biomech* 27(10):1201–1211
18. Liu MQ, Anderson FC, Pandy MG, Delp SL (2006) Muscles that support the body also modulate forward progression during walking. *J Biomech* 39(14):2623–2630
19. Erdemir A, McLean S, Herzog W, van den Bogert AJ (2007) Model-based estimation of muscle forces exerted during movements. *Clin Biomech (Bristol Avon)* 22(2):131–154
20. Pandy MG, Andriacchi TP (2010) Muscle and joint function in human locomotion. *Annu Rev Biomed Eng* 12:401–433
21. Viceconti M, Taddei F, Cristofolini L, Martelli S, Falcinelli C, Schileo E (2012) Are spontaneous fractures possible? An example of clinical application for personalised multiscale neuro-musculo-skeletal modelling. *J Biomech* 45(3):421–426
22. Bitsakos C, Kerner J, Fisher I, Amis AA (2005) The effect of muscle loading on the simulation of bone remodelling in the proximal femur. *J Biomech* 38(1):133–139
23. Claes L, Fiedler S, Ohnmacht M, Duda GN (2000) Initial stability of fully and partially cemented femoral stems. *Clin Biomech* 15(10):750–755
24. Ohman C, Dall'ara E, Baleani M, Van Sint Jan S, Viceconti M (2008) The effects of embalming using a 4 % formalin solution on the compressive mechanical properties of human cortical bone. *Clin Biomech (Bristol Avon)* 23(10):1294–1298
25. Yosibash Z, Tal D, Trabelsi N (2010) Predicting the yield of the proximal femur using high-order finite-element analysis with inhomogeneous orthotropic material properties. *Philos Transact Math Phys Eng Sci* 368(1920):2707–2723
26. Skurla CP, Pluhar GE, Frankel DJ, Egger EL, James SP (2005) Assessing the dog as a model for human total hip replacement analysis of 38 canine cemented femoral components retrieved at post-mortem. *J Bone Jt Surg Br* 87(1):120–127
27. Cristofolini L, Viceconti M (2000) Development and validation of a technique for strain measurement inside polymethyl methacrylate. *J Strain Anal Eng Des* 35(1):21–33
28. Sutton MA (2008) Editorial: special issue on image correlation methods. *J Strain Anal Eng Des* 43(8):1–2
29. Juszczczyk M, Pallini F, Schileo E, Taddei F, Cristofolini L (2006) Improvement of experimental-numerical cross-validation in studies of the proximal femur. *J Biomech* 39(S1):S473
30. Popper K (2003) *The logic of scientific discovery*. Taylor and Francis Publishing group, New York
31. Cristofolini L, Schileo E, Juszczczyk M, Taddei F, Martelli S, Viceconti M (2010) Mechanical testing of bones: the positive synergy of Fe models and in vitro experiments. *Philos Trans Math Phys Eng Sci* 368:2725–2763
32. Henninger HB, Reese SP, Anderson AE, Weiss JA (2010) Validation of computational models in biomechanics. *Proc Inst Mech Eng H* 224(7):801–812
33. Bedzinski R, Scigala K (2010) Biomechanical basis of tissue implant interactions. In: Kuczma M, Wilmanski K (eds) *Computer methods in mechanisms*, vol 1(1). Springer, New York, pp 379–390
34. Trabelsi N, Yosibash Z, Wutte C, Augat P, Eberle S (2011) Patient-specific finite element analysis of the human femur—a double-blinded biomechanical validation. *J Biomech* 44(9):1666–1672

35. Cristofolini L, Angeli E, Juszczak JM, Juszczak MM (2013) Shape and function of the diaphysis of the human tibia. *J Biomech* 46(11):1882–1892
36. Cristofolini L, Cappello A, Toni A (1994) Experimental errors in the application of photoelastic coatings on human femurs with uncemented hip stems. *Strain* 30(3):95–103
37. Cristofolini L, Metti C, Viceconti M (2003) Stress patterns induced by press-fitting and by external loads in cementless hip arthroplasty measured with photoelastic coatings. *J Strain Anal Eng Des* 38(4):289–301
38. Spirito G, Cristofolini L, Freddi A (1999) Deformation measurement by means of holographic interferometry with digital data analysis. In: Proceedings of abstracts XVI Danubia-Adria symposium on experimental methods in solid mechanics (Cluj-Napoca, Romania), pp 17–18
39. Merx H, Dreinhofer K, Schrader P, Sturmer T, Puhl W, Gunther KP, Brenner H (2003) International variation in hip replacement rates. *Ann Rheum Dis* 62(3):222–226
40. Cristofolini L, Viceconti M (2000) Mechanical validation of whole bone composite tibia models. *J Biomech* 33(3):279–288
41. Cristofolini L, Viceconti M, Cappello A, Toni A (1996) Mechanical validation of whole bone composite femur models. *J Biomech* 29(4):525–535
42. Carbone A, Howie DW, McGee M, Field J, Pearcy M, Smith N, Jones E (2006) Aging performance of a compliant layer bearing acetabular prosthesis in an ovine hip arthroplasty model. *J Arthroplast* 21(6):899–906
43. Bourne RB, Finlay JB (1986) The influence of tibial component intramedullary stems and implant-cortex contact on the strain distribution of the proximal tibia following total knee arthroplasty an in vitro study. *Clin Orthop Relat Res* 208:95–99
44. Little JP, Taddei F, Viceconti M, Murray DW, Gill HS (2007) Changes in femur stress after hip resurfacing arthroplasty: response to physiological loads. *Clin Biomech (Bristol Avon)* 22(4):440–448
45. Viceconti M, Affatato S, Baleani M, Bordini B, Cristofolini L, Taddei F (2009) Pre-clinical validation of joint prostheses: a systematic approach. *J Mech Behav Biomed Mater* 2(2):120–127
46. Viceconti M, Cristofolini L, Baleani M, Toni A (2001) Pre-clinical validation of a new partially cemented femoral prosthesis by synergetic use of numerical and experimental methods. *J Biomech* 34(6):723–731
47. Birolini A (2014) Reliability engineering, theory and practice, 7th edn. Springer, Berlin
48. Bowman NK, Bucher TA, Bassily AA (2006) Fracture of the stem of the femoral component after resurfacing arthroplasty of the hip. *J Bone Jt Surg Br* 88(12):1652–1653
49. Viceconti M, Ruggeri O, Toni A, Giunti A (1996) Design-related fretting wear in modular neck hip prosthesis. *J Biomed Mater Res* 30(2):181–186
50. Viceconti M, Baleani M, Squarzone S, Toni A (1997) Fretting wear in a modular neck hip prosthesis. *J Biomed Mater Res* 35(2):207–216
51. Currey JD (2001) Bone strength: what are we trying to measure? *Calcif Tissue Int* 68(4):205–210
52. Lengsfeld M, Burchard R, Gunther D, Pressel T, Schmitt J, Leppek R, Griss P (2005) Femoral strain changes after total hip arthroplasty-patient-specific finite element analyses 12 years after operation. *Med Eng Phys* 27(8):649–654
53. Speirs AD, Heller MO, Taylor WR, Duda GN, Perka C (2007) Influence of changes in stem positioning on femoral loading after THR using a short-stemmed hip implant. *Clin Biomech (Bristol Avon)* 22(4):431–439
54. Taddei F, Schileo E, Helgason B, Cristofolini L, Viceconti M (2007) The material mapping strategy influences the accuracy of CT-based finite element models of bones: an evaluation against experimental measurements. *Med Eng Phys* 29(9):973–979
55. Schileo E, Taddei F, Malandrino A, Cristofolini L, Viceconti M (2007) Subject-specific finite element models can accurately predict strain levels in long bones. *J Biomech* 40(13):2982–2989
56. Bayraktar HH, Gupta A, Kwon RY, Papadopoulos P, Keaveny TM (2004) The modified super-ellipsoid yield criterion for human trabecular bone. *J Biomech Eng* 126(6):677–684

57. Schileo E, Taddei F, Cristofolini L, Viceconti M (2008) Subject-specific finite element models implementing a maximum principal strain criterion are able to estimate failure risk and fracture location on human femurs tested in vitro. *J Biomech* 41(2):356–367
58. Cristofolini L, Erani P, Savigni P, Bordini B, Viceconti M (2007) Preclinical assessment of the long-term endurance of cemented hip stems. Part 2: in-vitro and ex-vivo fatigue damage of the cement mantle. *Proc Inst Mech Eng [H]* 221(6):585–599
59. Cristofolini L, Saponara Teutonico A, Savigni P, Erani P, Viceconti M (2007) Preclinical assessment of the long-term endurance of cemented hip stems. Part 1: effect of daily activities—a comparison of two load histories. *Proc Inst Mech Eng [H]* 221(6):569–584

Chapter 11

Reliability Models Based on Experiments

Giorgio Olmi

Abstract An engineered system may be affected by several uncertainties, which can be classified as *aleatory* or *epistemic* ones. To be able to manage the effects of these uncertainties is nowadays more and more important, especially in those cases where an accident may have catastrophic effects and it is necessary to predict its probability of occurrence. This is the case, for instance, of life extension problems, where the final decision must be supported by a probabilistic approach, to make sure that the probability of a fatality is sufficiently low. Experimentation is of course also important, as it is the way to estimate or reduce the uncertainty affecting the studied system. This chapter tackles the aforementioned issues, from the estimation of statistical distributions, based on experimental data, to the adoption of a *Most Probable Point Method* to estimate the probability of failure of engineered systems, depending on several variables, in an efficient and accurate way. Theory is accompanied by several practical case studies and exercises taken from existing research, which confirm the strict relationship between the outcomes of experimental campaigns and the development of analytical models for probability of failure prediction.

11.1 Introduction

Let us consider a quite recurrent and easy problem in the practice of a mechanical engineer: the structural analysis of a rod under a tensile load (see Fig. 11.1a). First of all, the axial load N should be determined, considering the forces acting on the entire structure and the amount of load being transmitted to the rod. Then, a suitable cross section should be worked out, let the term A denote its area. The rod should be verified, computing the stress $\sigma = \frac{N}{A}$, to be compared to the resistance. If only static loads are acting on the structure, then the static (ultimate or yield, depending on the application) strength, determined by a tensile test on a specimen of the same material, should be considered. Otherwise, if the loads experience a cyclic variation with millions of loops expected over the rod life, the fatigue limit should be accounted for. The structure is therefore verified, if σ is lower than the related static or fatigue strength. A safety coefficient is commonly introduced, to quantify the degree of safety, i.e. how greater the strength is with respect to the expected stress level. The

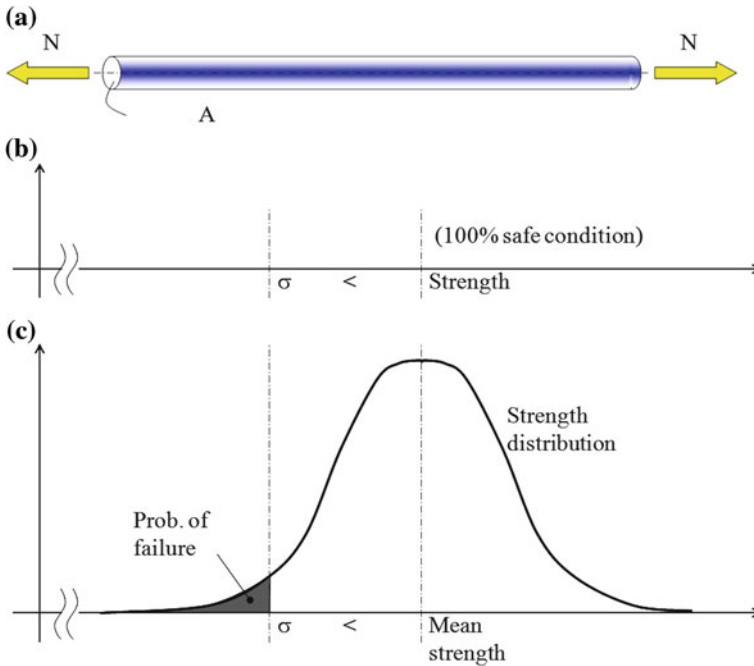


Fig. 11.1 a A sketch of the rod under tension to be verified, b *deterministic* approach, c *probabilistic* approach and probability of failure (*highlighted area*)

same procedure can be followed when a multi-axial stress state is generated in a component. In this case, an equivalent stress must be computed and compared with the corresponding strength. What has been described is the typical *deterministic* approach, where stress and strength are regarded as constants, and the structure is expected to be safe or to collapse, respectively when the stress level is lower or greater than strength.

What can be remarked is that an engineered system may be affected by several uncertainties. In Refs. [1, 2] it is pointed out that uncertainty can be viewed as the difference between the present state of knowledge and the complete knowledge and is usually classified into *aleatory* and *epistemic* types. The first one is due to the inherent variability of the system, whereas the second one is a consequence of a poor level of information. For instance, in the case of the rod, the load entity may have unexpected fluctuations during use. Sometimes, in view of the fact that a load history is transmitted to the rod, it may be difficult to achieve an accurate measurement of the load trend. The aforementioned uncertainties can be regarded as *epistemic*. A further issue regards strength: this strictly depends on the material and can be determined, in the case of a static load, by running a tensile test. Problems may arise from the occurrence that characteristics may be different, depending on the material lot. As a consequence, if the tensile test is repeated many times on different specimens, a

scattering of results is likely to be observed, which can be regarded as an *aleatory* uncertainty. In addition, material mechanical testing is indeed also affected by an inherent uncertainty, arising from the noise in the measuring chain for the detection of the applied load.

Coming back to our example, the final outcome is that both the load (and, consequently, the stress) and the strength must be conveniently regarded as random variables. Each of them has a distribution that can be estimated by in-field monitoring of loads and by the statistical analysis of the yields of a sufficiently large number of tensile tests. The two approaches of *deterministic* and *probabilistic* verification of the rod are compared in Fig. 11.1b, c; in this last case, a deterministic load and a randomly normally distributed strength are considered.

Being able to compute a probability of failure, indicated by the highlighted area in Fig. 11.1c, makes it possible to manage the effects of uncertainty and to develop engineering solutions that take random fluctuations of variables into account [3–17]. This is particularly important in all those applications, where the variability of factors is significant or/where failures may result in catastrophic effects [10]. In all these cases a probability approach is required, which enables the comparison between the determined probability of failure and reference values, or acceptability thresholds, depending on the severity of the effects. Practical examples are in the fields of aerospace, of nuclear power plants, and of structures under low- and high-cycle fatigue. Highly difficult problems in the practice of the engineer, such as the life extension of a largely scaled and expensive machine, for instance a turbine, a pressure vessel, a plane cockpit, are more conveniently tackled from a probabilistic point of view. A life extension decision must be supported by a probability of failure being below an acceptability threshold.

The main questions that are going to be tackled in this chapter are summarized in the following points.

- How to determine the statistical distribution of a random variable, based on experimentally sampled values.
- How to process experimental data in regression lines and how to account for their scattering.
- How to determine a likelihood band of a regression and how to assign a reasonable statistical distributions to its slope and offset terms.
- How to determine the distribution of an output variable, being related to a number of input variables, in the general case of an implicit relationship.
- How to conduct reliability assessments in common case studies in the engineering field, determining a probability of failure supported by experimental data.

11.2 Determination of the Distributions of Sets of Data

This Section is organized in three Case Studies, developed in the next Subsections. The following questions are tackled:

- Estimation of a statistical distribution, based on a finite set of data.
- Determination of the probability of failure when one single random variable is affecting the output.
- Quality control.

A further example dealing with industrial quality improvement is provided in [18].

11.2.1 Case Study: Height Distribution in a Class of Students

This first Case Study introduces to the general procedure to arrange a set of data in a histogram and then to work out the related statistical distribution.

11.2.1.1 Introduction to the Problem

The data in Table 11.1 refer to the heights of a population of 80 students. How is it possible to convert these data into a distribution? How is it possible to test the adequacy of the distribution in the representation of the data?

11.2.1.2 Histograms for the Graphical Representation of the Results

The first remark to be made is related to the size of the population. In order to obtain a statistically significant distribution, estimated on the basis of a sufficiently sized finite number of yields, it should not be lower than 50 [19, 20]. The first step in the processing of the data consists in their discretization: in other words, the entire population, distributed over an overall range, called *excursion*, must be split into a discrete number of *categories*. The first issue lies in the determination of the most suitable number of categories (*k*). According to many References (e.g., [19, 20]), it

Table 11.1 Heights of a population of 80 students

Sample number	Height (m)									
1–10	1.83	1.85	1.49	1.85	1.71	1.46	1.61	1.68	1.91	1.92
11–20	1.55	2.00	1.86	1.67	1.81	1.51	1.66	1.85	1.77	1.91
21–30	1.74	1.39	1.85	1.85	1.75	1.75	1.75	1.66	1.72	1.55
31–40	1.75	1.31	1.61	1.40	1.46	1.85	1.75	1.62	1.86	1.34
41–50	1.66	1.66	1.75	1.79	1.55	1.67	1.66	1.72	1.75	1.75
51–60	1.61	1.75	1.72	1.55	1.49	1.67	1.91	1.65	1.71	1.56
61–70	1.75	1.60	1.68	1.75	1.85	1.91	1.68	1.50	1.52	1.61
71–80	1.85	1.59	1.82	1.56	1.85	1.66	1.55	1.57	1.71	1.66

can be computed as in Eq. 11.1, where the symbol N stands for the overall number of the data and \log indicates a 10-base logarithm.

$$k = 1 + 3.3 \cdot \log(N) \tag{11.1}$$

In the present case, for $N = 80$, we have $k = 7.2$, to be rounded to 7. The most suitable value yielded by Eq. 11.1 must be intended as a compromise between a too low value, which would lead to a rough discretization, and a too high one. In this last case, the processing of the data would be made ineffective. With the data split into a high number of classes, it would be impossible to find the interval where the results are more concentrated and therefore to get an interpretation of the distribution. The size of each category can be easily determined as the quotient between the *excursion* and the determined number of categories, considering uniformly sized categories. In this case, we obtain 0.10 m-wide categories.

Therefore, the results can be collected in a table, like the one in Table 11.2, and graphically represented, as in the histogram in Fig. 11.2. The *occurrence* n_i stands for the number of students, whose heights are within each *category*. It is clear that $\sum_{i=1}^k n_i = N$.

The following step consists in the computation of the *frequencies* f_i of the results, easily determined as the ratios between the *occurrences* and the total number of results: $f_i = \frac{n_i}{N}$. It is clear that $\sum_{i=1}^k f_i = \sum_{i=1}^k \frac{n_i}{N} = 1$. The results in terms of *frequencies* are summarized in the histogram in Fig. 11.3.

The last step consists in the computation of category *densities* δ_i . Each *density* is given by the ratio between the related *frequency* (f_i) and the category size (d):

Table 11.2 Categories: ranges and occurrences of results

	Categories						
Ranges (m)	1.30–1.40	1.40–1.50	1.50–1.60	1.60–1.70	1.70–1.80	1.80–1.90	1.90–2.00
Mean values (m)	1.35	1.45	1.55	1.65	1.75	1.85	1.95
Occurrences (n_i)	4	5	12	19	20	14	6

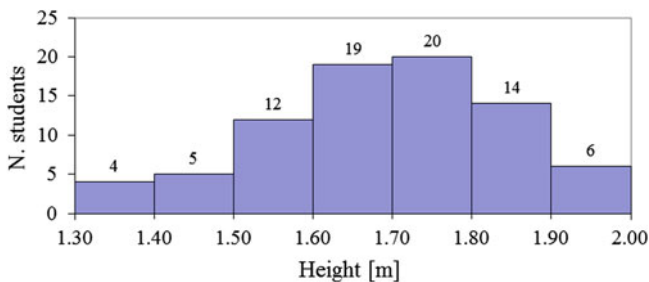


Fig. 11.2 Height measurements split into the seven categories and related *occurrences*

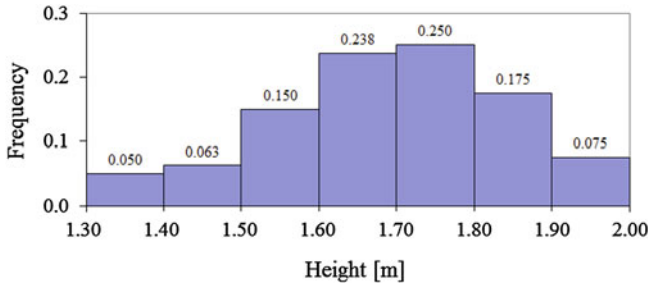


Fig. 11.3 Data frequencies over the seven categories

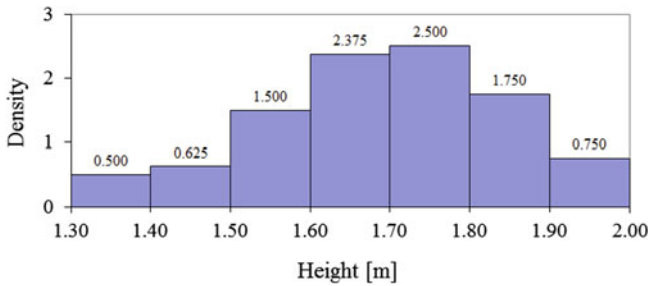


Fig. 11.4 Data densities over the seven categories

$\delta_i = \frac{f_i}{d}$ ($\delta_i = \frac{f_i}{d_i}$, should the categories have different sizes). The histogram in terms of densities is shown in Fig. 11.4.

Considering this last graph, it can be easily shown that the sum of the areas of the bars is unitarian. The total area A is yielded by $\sum_{i=1}^k (d \cdot \delta_i) = \sum_{i=1}^k (d \cdot \frac{f_i}{d}) = 1$. Therefore, the graph in Fig. 11.4 may be regarded as a discrete representation of the distribution.

11.2.1.3 Normal Probability Plot

A careful examination of the histogram in Fig. 11.4 suggests that the data are normally distributed around their mean value. However, hastily concluding that they are normally distributed could be questionable at this stage. An important issue is therefore a check of the adequacy of the normal model. For this purpose, the tool of the *normal probability plot*, and in particular of the *quantile-quantile plot*, can be used, to overcome the aforementioned question. In order to apply this tool, the data in Table 11.1 must be arranged in increasing order. Afterwards, it is necessary to compute the *standardized normal quantiles*, to be related one-by-one to the ordered data. Generally considering N yields to be processed, the *first standardized normal quantile* can be defined as value Z_1 that corresponds, in a *standard normal*

distribution, to a cumulated area (from $-\infty$ to Z_1), whose extension is $\frac{1}{1+N}$. In the same way, the j th quantile is the value Z_j along the horizontal axis of a standard normal distribution, corresponding to a cumulated area of $\frac{j}{1+N}$. Finally, the N th quantile (the last one) is the Z_N value for a cumulated area, equating $\frac{N}{1+N}$, i.e. a bit lower than one, corresponding to the cumulated area of the full distribution. In short, the j th quantile is yielded by Eq. 11.2, where ϕ stands for the standard normal cumulative distribution.

$$Z_j = \phi^{-1} \left(\frac{j}{1+N} \right) \tag{11.2}$$

Some results, obtained by the application of Eq. 11.2, are summarized in Table 11.3.

The normal probability plot is finally obtained plotting the experimental data versus the corresponding quantiles. In particular, the yield for $j = 1$ (i.e.: the lowest one in the ordered list) is plotted versus the quantile Z_1 , and so on. Considering the present application, the described procedure is implemented in the diagram in Fig. 11.5, where the ordered heights are plotted versus the Z terms, referring to the rows of Table 11.3. The analysis of the resulting graph suggests that the points follow a linear trend, as confirmed also by the high linear correlation coefficient R^2 being very close to 1. A linear trend indicates that a normal distribution adequately describes the data population. [21]. Should the trend be non-linear, asymmetric distributions (generally, Weibull distributions) must be considered for the processing of the yields [20–22].

Table 11.3 Computation of quantiles

Sample number (j)	Height (m)	Cumulative area $\left(\frac{j}{1+N}\right)$	Z_j
1	1.31	0.0123	-2.246
2	1.34	0.0247	-1.965
3	1.39	0.0370	-1.786
4	1.40	0.0494	-1.651
5	1.46	0.0617	-1.540
...
76	1.91	0.9383	1.540
77	1.91	0.95063	1.651
78	1.91	0.9630	1.786
79	1.92	0.9753	1.965
80	2.00	0.9877	2.246

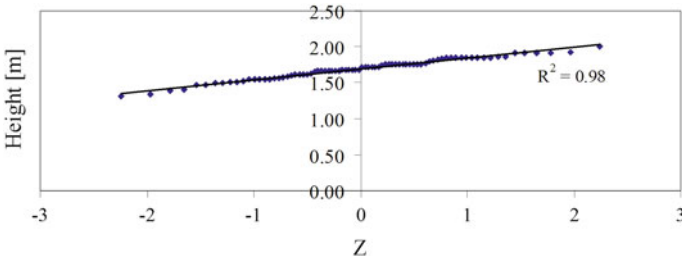


Fig. 11.5 Normal probability plot, where the data in increasing order are plotted versus the quantiles

11.2.1.4 Distributions

Once it has been asserted that the data are normally distributed, this distribution is univocally defined by its mean and standard deviation. Reliable estimations of these terms can be retrieved by the computation of the mean and of the standard deviation of the finite number of yields in Table 11.1. The results in terms of the *probability density function (pdf)* and of the *cumulative distribution function (cdf)* are shown in Fig. 11.6.

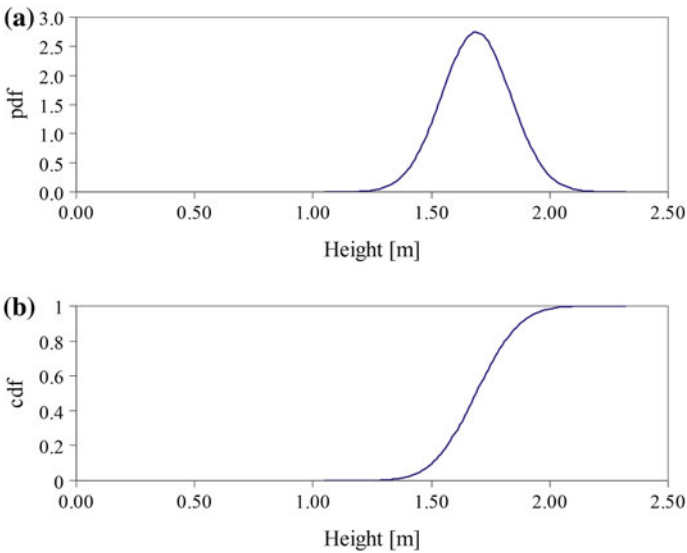


Fig. 11.6 a Retrieved probability density function (pdf) and b cumulative distribution function (cdf)

11.2.2 Case Study: Probability of Static Failure of a Tensioned Rod

This second Case Study is a first approach to the probability of failure in a structural problem.

11.2.2.1 Introduction to the Problem

A rod, having a circular cross section with diameter $d = 26$ mm, is loaded by an axial static force of $F = 120$ kN. The material has been characterized, by running 60 tensile tests, according to the Standard [23]. The yields are listed (in the trial order) in Table 11.4. The problem consists in determining the distribution of the static strength of the bar material and the probability of failure of the rod under the aforementioned load.

The first issue to be considered consists in the size of the finite population to be considered to infer the statistical distribution. The number of 60 samples can be regarded as large enough to determine the distribution. The steps described in Sect. 11.2.1 must be followed, to process the data in the form of histograms, to test the adequacy of the distribution model and to determine the distribution itself.

11.2.2.2 Histograms for the Graphical Representation of the Results

The results range from approximately 210–290 MPa with an overall range of 80 MPa. This range must be split into a suitable number of categories, determined by Eq. 11.1. The result for $N = 60$, 6.87, suggests that seven categories can be properly utilized. Thus, considering uniformly sized classes, a width of about 11.4 MPa can be easily obtained. The yields are collected and grouped into the seven categories in Table 11.5.

The related histogram, with the *occurrences* split into the seven *categories*, is shown in Fig. 11.7. The following steps consist in the calculation of the *frequencies*

Table 11.4 Strengths of a population of 60 specimens

Sample number	Strength (MPa)									
1–10	249	262	253	251	268	245	255	254	249	253
11–20	226	225	251	257	270	235	253	251	223	249
21–30	241	257	247	251	242	253	245	254	254	254
31–40	251	233	244	267	240	262	251	286	232	250
41–50	235	275	212	256	262	264	233	249	245	260
51–60	249	266	243	245	236	236	265	253	252	236

Table 11.5 Categories: ranges and occurrences of results

	Categories						
Ranges (MPa)	210.0– 221.4	221.4– 232.9	232.9– 244.3	244.3– 255.7	255.7– 267.2	267.2– 278.6	278.6– 290.0
Mean values (MPa)	215.7	227.2	238.6	250.0	261.4	272.9	284.3
Occurrences (n_i)	1	4	12	28	11	3	1

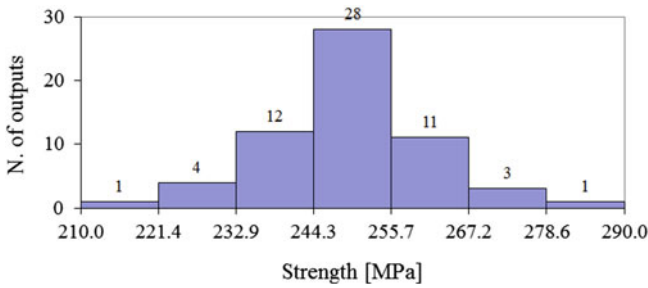


Fig. 11.7 Determined tensile strengths split into the seven categories and related occurrences

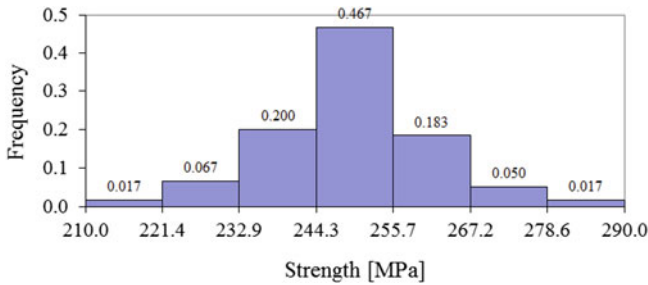


Fig. 11.8 frequencies of the strength results over the seven categories

$f_i = \frac{n_i}{N}$ and of the densities $\delta_i = \frac{f_i}{d}$, where d indicates the uniform category size. The related histograms are plotted in Figs. 11.8 and 11.9.

11.2.2.3 Normal Probability Plot

In this case too, the results appear to be well distributed around their central value. The adequacy of the normal distribution can be verified by the tool of the *quantile-quantile plot*. For this purpose, the data must be rearranged as in Table 11.6, where the columns refer to the data arranged in increasing order, to the probability terms $\left(\frac{j}{1+N}\right)$ and to the Z_j terms, corresponding to the cumulated probabilities in a *standard normal distribution* ϕ (see also Eq. 11.2).

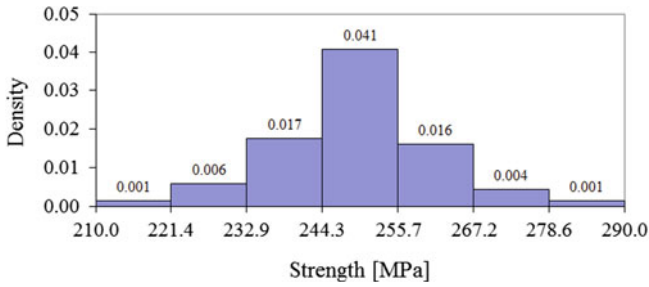


Fig. 11.9 Data densities over the seven categories

Table 11.6 Computation of quantiles

Sample number (j)	Strength (MPa)	Cumulative area $\left(\frac{j}{1+N}\right)$	Z_j
1	212	0.0164	-2.1347
2	223	0.0328	-1.8413
3	225	0.0492	-1.6529
4	226	0.0656	-1.5096
5	232	0.0820	-1.3920
...
56	267	0.9180	1.3920
57	268	0.9344	1.5096
58	270	0.9508	1.6529
59	275	0.9672	1.8413
60	286	0.9836	2.1347

The normal probability plot is shown in Fig. 11.10, where the yields of the material characterization are plotted versus the quantiles Z indicated in Table 11.6. The data are included in the graph, maintaining their increasing order, so that the j th value

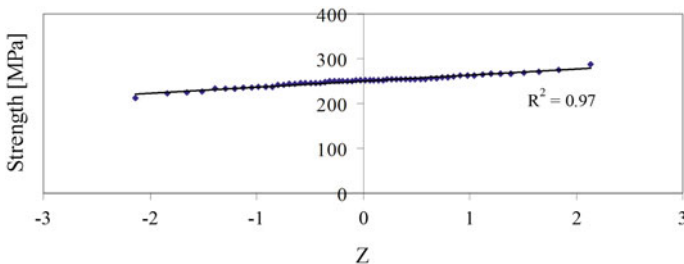


Fig. 11.10 Normal probability plot: strengths in increasing order plotted versus the Z_j quantiles

of strength (corresponding to the j th row of Table 11.6) is plotted versus Z_j . The dots are well aligned along a straight line, as confirmed by the high linear correlation coefficient $R^2 = 0.97$. Therefore, a normal distribution can be correctly assumed to describe the population of data.

11.2.2.4 Distributions and Probability Estimation

The mean and the standard deviation of the normal distribution can be reasonably estimated as the mean and the standard deviation of the finite population of yields, respectively 249.4 and 13.12 MPa. The related distributions, *pdf* and *cdf*, are plotted in Fig. 11.11.

Coming back to the initial question, the tensile stress σ on the rod, uniformly distributed over its circular cross section, can be easily determined as in Eq. 11.3.

$$\sigma = \frac{F}{\frac{\pi \cdot d^2}{4}} = \frac{120 \times 10^3}{\frac{\pi \cdot 26^2}{4}} = 226 \text{ MPa} \tag{11.3}$$

Failure occurs when the strength is lower than the computed stress, therefore the probability of failure p_f is given by the highlighted area, i.e. the tail of the *pdf*

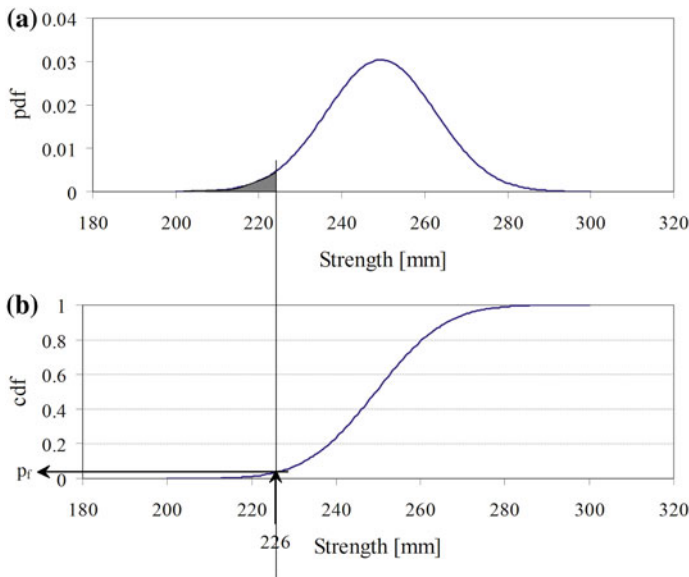


Fig. 11.11 **a** Retrieved *probability density function (pdf)* and **b** *cumulative distribution function (cdf)*, determination of p_f

or the cumulated probability yielded by the *cdf*, entering the retrieved value of σ . Considering the distributions in Fig. 11.11, p_f is approximately 3.7 %.

11.2.3 Case Study: Quality Control of Pin Diameters

This last Case Study shows how the described procedure can be used as a support to Quality Control in industrial applications.

11.2.3.1 Introduction to the Problem

Pins with a nominal diameter of 30 mm are manufactured to be coupled to hubs. A population consisting of 50 samples is extracted to perform a dimensional control on the entity of the coupling diameters. The results are collected in Table 11.7. Considering that the tolerance band is between 29.8 and 30.2 mm, it is necessary to determine the probability of pins being not conformal with the design specifics and to indicate a possible strategy to improve the quality of the process, reducing the rejection rate.

The first issue to be considered consists in the size of the finite population of data to be processed to determine the statistical distribution. A number of 50 samples can be regarded as the minimum one to estimate a probability distribution with a sufficient statistical relevance. The steps in Sects. 11.2.1 and 11.2.2 are followed to determine the related distributions.

11.2.3.2 Histograms for the Graphical Representation of the Results

The results range from approximately 29.75–30.10 mm with an overall range of 0.25 mm. This range must be split into a suitable number of categories. According to Eq. 11.1 this number is between 6 and 7, therefore seven categories can be conveniently utilized. Considering uniformly sized classes, their width is 0.05 mm. This

Table 11.7 Coupling diameters of a population of 50 samples

Sample number	Diameter (mm)									
1–10	30.00	29.93	29.89	29.93	29.90	29.80	29.88	29.84	29.88	29.89
11–20	29.91	29.79	30.03	30.04	29.92	29.91	29.89	30.02	29.90	29.82
21–30	29.97	29.90	29.89	29.90	29.82	29.86	30.04	30.07	29.93	29.80
31–40	29.88	29.89	29.97	29.77	29.78	29.86	29.94	29.95	29.94	29.91
41–50	29.93	29.89	29.96	29.88	29.94	29.87	29.90	29.94	29.96	29.82

Table 11.8 Categories: ranges and occurrences of results

	Categories						
Ranges (mm)	29.75– 29.80	29.80– 29.85	29.85– 29.90	29.90– 29.95	29.95– 30.00	30.00– 30.05	30.05– 30.10
Mean values (mm)	29.78	29.83	29.88	29.93	29.98	30.03	30.08
Occurrences (n_i)	5	4	18	13	5	4	1

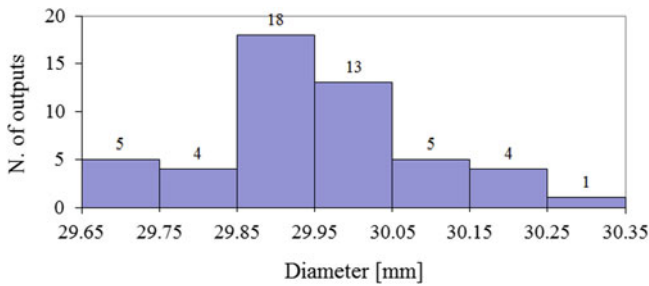


Fig. 11.12 Measured diameters split into the seven categories: related occurrences

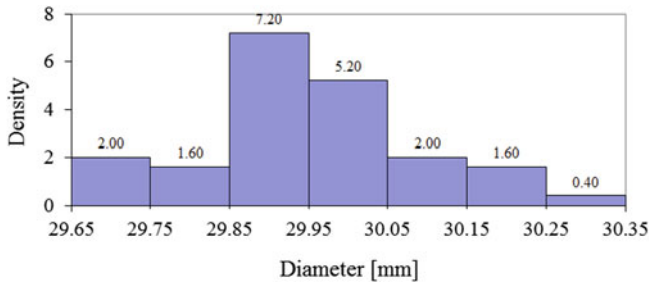


Fig. 11.13 Data densities over the seven categories

leads to the discretization in Table 11.8, where the measurements are grouped into the seven categories.

The related histogram is shown in Fig. 11.12. Afterwards, the occurrences n_i have been converted into the related frequencies $f_i = \frac{n_i}{N}$ and into the densities $\delta_i = \frac{f_i}{d}$, where d stands for the uniform width of each class. The histogram with densities plotted versus the category ranges is shown in Fig. 11.13.

11.2.3.3 Normal Probability Plot

It is interesting to remark that in the present case there is no clear evidence that the results are symmetrically distributed around their central value. Therefore, it is very important to check the adequacy of the normal distribution model by the tool of

the *quantile-quantile plot*. Should it not be suitable, more complicated distributions, such as the *Weibull* distribution, should be used to account for the non symmetric shape. In order to arrange the data in the *normal probability plot*, they have been ordered in the second column of Table 11.9. The values of the cumulated probability $\left(\frac{j}{1+N}\right)$ and of the *quantiles* Z_j are appended in the remaining columns of the same Table.

The *normal probability plot* is shown in Fig. 11.14, where the sample diameters are plotted versus the *quantiles* in Table 11.9. The results appear to follow a linear trend, with a linear correlation coefficient $R^2 = 0.97$. Therefore, the hypothesis of a normal distribution can be accepted.

Table 11.9 Computation of *quantiles*

Sample number (j)	Diameter (mm)	Cumulative area $\left(\frac{j}{1+N}\right)$	Z_j
1	29.77	0.0196	-2.0619
2	29.78	0.0392	-1.7599
3	29.79	0.0588	-1.5647
4	29.80	0.0784	-1.4157
5	29.80	0.0980	-1.2928
...
46	30.02	0.9020	1.2928
47	30.03	0.9216	1.4157
48	30.04	0.9412	1.5647
49	30.04	0.9608	1.7599
50	30.07	0.9804	2.0619

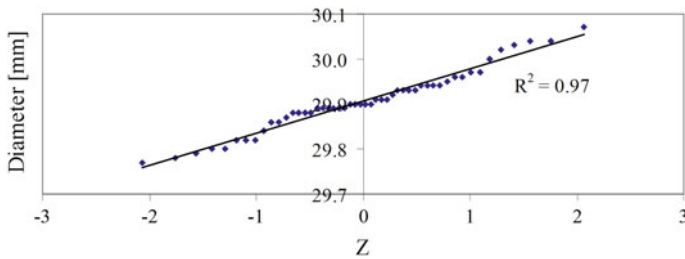


Fig. 11.14 *Normal probability plot*, where the retrieved diameters are plotted versus the related *quantiles*

11.2.3.4 Distributions and Probability Estimation

The mean and the standard deviation of the normal distribution can be reasonably estimated as the mean and the standard deviation of the 50-sized population of data, respectively 29.91 and 0.069 mm. The related distributions, *pdf* and *cdf*, can therefore be plotted in Fig. 11.15.

Coming back to the initial question, the probability that the manufactured pins are out of range (i.e.: with a diameter lower than 29.8 mm or higher than 30.2 mm) is given by the complement to one of the highlighted area in the *pdf* in Fig. 11.15a. The probability of part rejection is therefore around 6 %, which is quite high for an industrial process.

The analysis of the obtained distribution leads to the remark that the pins are generally rejected, due to their low diameter, being below the threshold of 29.8 mm. On the other hand, the maximum threshold of 30.2 mm is generally accomplished. Moreover, it is interesting to observe that the mean value, 29.91 mm, is a long way from the nominal dimension of 30 mm. Therefore, the quality can be strongly improved, if the manufacturing process can be modified to increase the pin diameter, so that its mean is made comparable to the nominal dimension. Supposing that the same standard deviation, which can be regarded as acceptable, can be maintained, this improvement leads to a shift of the *pdf*, as in Fig. 11.16. The probability of non-conformity in this case, again computed as the complement to one of the highlighted area, is significantly reduced to the value of 0.35 %.

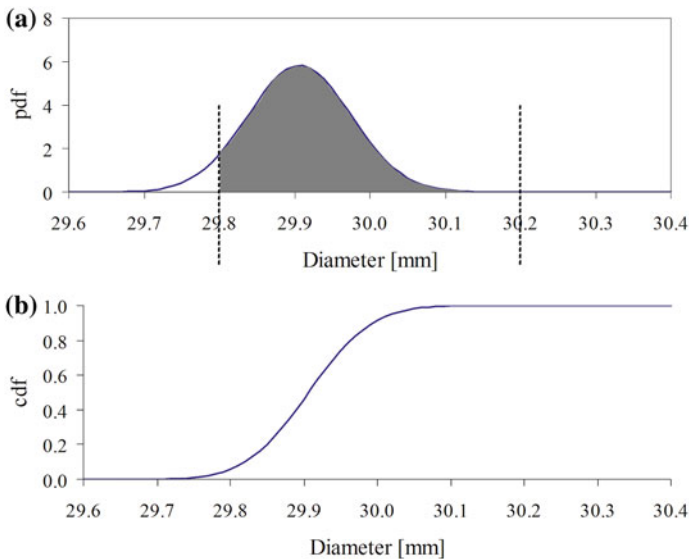


Fig. 11.15 **a** Retrieved *Probability density function (pdf)* with highlighted area corresponding to the complement to one of the defect rate and **b** *cumulative distribution function (cdf)*

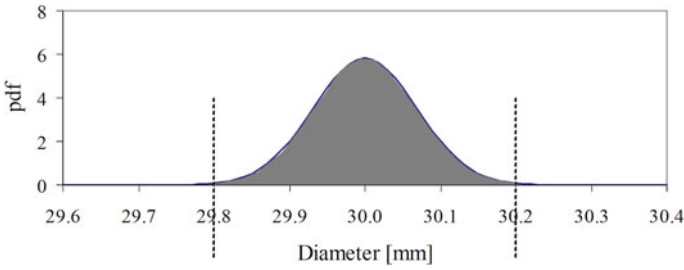


Fig. 11.16 New *Probability density function (pdf)* in the case of process improvement with highlighted area corresponding to the complement to one of the defect rate

11.3 How to Linearly Interpolate a Set of Data and How to Account for Their Scatter

Linearly interpolating a set of points is quite a common practice in experimental mechanics and basic statistics. The question can be easily tackled and assisted by many software and spread sheets. For instance, let x_i and y_i , Table 11.10, be the coordinates of a set of points, where x stands for the independent variable (input) and y indicates the dependent one (output).

The regression line is in the form $y = mx + q$, where the slope m and the offset term q can be easily determined by Eqs. 11.4 and 11.5, according to the mean-square approach.

Table 11.10 Points to be linearly interpolated

Sample number (i)	x_i	y_i
1	0	3.675
2	2	4.023
3	4	5.140
4	6	5.300
5	8	5.156
6	10	5.958
7	12	5.890
8	14	7.154
9	16	7.352
10	18	7.210
11	20	8.545

$$m = \frac{\sum_{i=1}^N [(x_i - \bar{x}) \cdot (y_i - \bar{y})]}{\sum_{i=1}^N (x_i - \bar{x})^2} \tag{11.4}$$

$$q = \bar{y} - m \cdot \bar{x} \tag{11.5}$$

The values of \bar{x} and \bar{y} (averages) are yielded by Eqs. 11.6 and 11.7.

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} \tag{11.6}$$

$$\bar{y} = \frac{\sum_{i=1}^N y_i}{N} \tag{11.7}$$

The result of the interpolation procedure is shown in Fig. 11.17, where the regression line is sketched in a solid line. An important issue is represented by the scattering of the data points, distributed around the determined regression line. This is crucial, when the experimental data are used to process an analytical law, describing the response of a device, for instance a load cell, or that of a material under a cyclic load, for instance the load-life relationship. Especially in this last case, the use of the plain regression line may result in an uncertain estimation, and sometimes in an overestimation of the expected life, with unsafe dimensioning in design applications. Therefore, it is important that the inferred model also accounts for the scattering of the experimental evidence. A possible option to overcome the problem is the determination of a maximum likelihood band, wrapped around the regression line and enclosed between two bounds. For instance, regarding fatigue, the procedure for

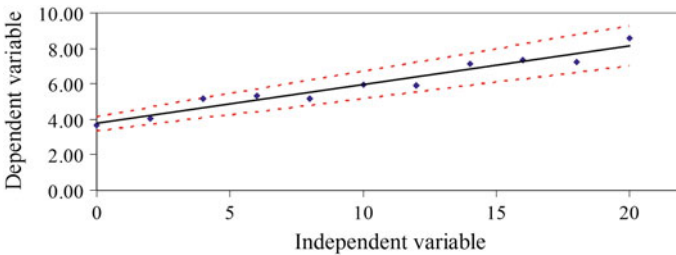


Fig. 11.17 Interpolation of a set of points, with related *lower* and *upper* bounds

the determination of the aforementioned bounds is provided in some Standards and papers [24–26].

In the described application, the following approach can be used for the determination, at a first stage, of the standard deviations of the slope (m) and of the offset term (q), respectively σ_m and σ_q . First of all, based also on [27], the terms S_x , S_y , S_{xx} , S_{xy} and Δ must be computed, according to Eqs. 11.8–11.12.

$$S_x = \sum_{i=1}^N x_i \quad (11.8)$$

$$S_y = \sum_{i=1}^N y_i \quad (11.9)$$

$$S_{xx} = \sum_{i=1}^N x_i^2 \quad (11.10)$$

$$S_{xy} = \sum_{i=1}^N (x_i \cdot y_i) \quad (11.11)$$

$$\Delta = S \cdot S_{xx} - S_x^2 \quad (11.12)$$

Secondly, a standard deviation that accounts for the scattering of the dots with respect to the determined line, must be computed. This term, named $\bar{\sigma}$ is yielded by Eq. 11.13, where \hat{y}_i stands for the estimation of the value of the dependent variable y_i for x_i , according to the determined regression line. It can be easily observed that the term at the numerator stands for the sum of the squares of the residuals, whereas the denominator contains the number of degrees of freedom.

$$\bar{\sigma} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N - 2}} \quad (11.13)$$

Finally, the standard deviation to be applied to the slope, σ_m and that of the offset term, σ_q , are respectively yielded by Eqs. 11.14 and 11.15.

$$\sigma_m = \bar{\sigma} \cdot \sqrt{\frac{S}{\Delta}} \quad (11.14)$$

$$\sigma_q = \bar{\sigma} \cdot \sqrt{\frac{S_{xx}}{\Delta}} \quad (11.15)$$

The last step is the determination of the lower and upper bounds. Considering for instance a scenario of twice the standard deviation, the lower and upper values for m and q can be determined by Eqs. 11.16 and 11.17.

$$\begin{aligned} m_{min} &= m - 2 \cdot \sigma_m \\ m_{max} &= m + 2 \cdot \sigma_m \end{aligned} \quad (11.16)$$

$$\begin{aligned} q_{min} &= m - 2 \cdot \sigma_q \\ q_{max} &= m + 2 \cdot \sigma_q \end{aligned} \quad (11.17)$$

Finally, the lower and upper bounds, respectively in the form $y_{min} = m_{min}x + q_{min}$ and $y_{max} = m_{max}x + q_{max}$, are plotted in the dashed line in Fig. 11.17.

11.3.1 How to Determine the Coefficient of Plasticity and the Hardening Exponent of a Material and Related Distributions

A tensile test is usually performed according to the recommendations by the Standard [23] and operating in the displacement controlled mode. During the test, the strain and the stress at the specimen gage are measured respectively by an extensometer and by the testing machine load cell. Many models are available in literature for the analytical description of the obtained stress-strain curve, considering its elastic and plastic parts and related transition [28–31]. One of the most used is the model by Ramberg and Osgood [30], whose formulation is recalled in Eq. 11.18.

$$\varepsilon = \varepsilon_{el.} + \varepsilon_{pl.} = \frac{\sigma}{E} + \left(\frac{\sigma}{K}\right)^n \quad (11.18)$$

In the stress-strain ($\sigma - \varepsilon$) relationship of Eq. 11.18, the term E stands for the Young's modulus, whereas K and n are respectively the *static coefficient of plasticity* and the *hardening exponent*. They can be easily determined, separating the amounts of elastic ($\varepsilon_{el.}$) and plastic $\varepsilon_{pl.}$ strains. A linear regression in the logarithmic scale must be subsequently run on the plastic part of strain, as in Eq. 11.19.

$$\log(\sigma) = n \cdot \log(\varepsilon_{pl.}) + \log(K) \quad (11.19)$$

The terms n and $\log(K)$ are finally determined as the slope and the offset term of the interpolating straight line. Therefore, the procedure described in Sect. 11.3 and the formulations in Eqs. 11.14 and 11.15 can be applied for the computation of the standard deviations of these terms.

The question related to the distributions of the *static coefficient of plasticity* and of the *hardening exponent* must be properly tackled, considering first the most widely

accepted approaches in the literature. Regarding the *coefficient of plasticity* (both *static* and *cyclic*), it is generally [11, 12, 32–35] presumed to be log-normally distributed, meaning that its logarithm has a normal distribution. As far as the hardening exponent (both *static* and *cyclic*) is concerned, different options are considered in the existing studies. Sometimes it has been regarded as deterministic [36, 37], in one case a log-normal distribution was attributed [33], while in the other cases [11, 12, 34, 35] the normal distribution was indicated as the most suitable.

Therefore, in agreement with the literature, both n and $\log(K)$ can be supposed to be normally distributed: related distributions are univocally determined by their nominal values, which can be regarded as the means, and by the computed standard deviations. Finally, the Young's modulus E can be conveniently considered as a deterministic constant, as supported by several references [11, 12, 33, 36–38].

Exercise 11.1 (*Cyclic curve of a steel material: distributions of the cyclic coefficient of plasticity and of the hardening exponent*) The procedure described in Sect. 11.3.1 can also be applied to the determination of the cyclic curve of a material and related parameters. The following example considers this case.

A steel material ($E = 193.8$ GPa), with wide applications in the manufacturing of turbogenerator rotors, has been characterized under low cycle fatigue [26, 39, 40], according to the Standards [41, 42]. In particular, its cyclic curve has been determined as an envelope of the upper tips of steady-state stress-strain hysteresis loops. The experimental results, considering 25 loops, are summarized in Table 11.11.

The linear regression of the points in the two last columns of Table 11.11 leads to the computation of the slope and of the offset term, namely $n' = 0.052$ and $\log(K') = 2.949$. A superscript has been added to denote that these are the *cyclic* coefficient of plasticity and hardening exponent. According to the procedure in the present Section, the following terms can be easily calculated, based on the data in Table 11.11.

- $S = 24$
- $S_x = -63.89$
- $S_y = 67.44$
- $S_{xx} = 176.59$
- $S_{xy} = -179.19$
- $\Delta = 156.34$
- $\bar{\sigma} = 0.0091$

Finally, the previous results make it possible to calculate the standard deviations of the slope n' and of the offset $\log(K')$ are respectively 3.6×10^{-3} and 9.7×10^{-3} . This outcome makes it possible to determine the distributions of the parameters that characterize the cyclic behaviour of the studied material.

The interpolating line is shown in the graph in Fig. 11.18 in a $\log(\sigma) - \log(\varepsilon_{pl.})$ reference system, together with its bounds, for a scenario of twice the standard deviation. The determined cyclic curve is depicted in Fig. 11.19 along with the steady-state loops. Lower and upper bounds, wrapped around its plastic part, are added,

Table 11.11 Steady state stress-strain hysteresis loops

Strain (%)	Strain (-)	Elastic strain ($\varepsilon_{el.}$) (-)	Plastic strain ($\varepsilon_{pl.}$) (-)	Stress (σ (MPa))	x ($\log(\varepsilon_{pl.})$)	y ($\log(\sigma)$)
0.30	0.003	0.003	9.53×10^{-5}	563	-4.021	2.750
0.30	0.003	0.003	1.98×10^{-4}	543	-3.703	2.735
0.30	0.003	0.003	1.48×10^{-4}	553	-3.830	2.743
0.39	0.004	0.003	0.001	616	-3.141	2.789
0.42	0.004	0.003	0.001	630	-3.023	2.799
0.43	0.004	0.003	0.001	622	-2.962	2.794
0.51	0.005	0.003	0.002	650	-2.758	2.813
0.58	0.006	0.003	0.003	637	-2.600	2.804
0.59	0.006	0.003	0.003	655	-2.598	2.816
0.63	0.006	0.003	0.003	657	-2.536	2.818
0.62	0.006	0.003	0.003	648	-2.544	2.812
0.64	0.006	0.004	0.003	682	-2.540	2.834
0.65	0.007	0.003	0.003	663	-2.511	2.821
0.72	0.007	0.003	0.004	666	-2.425	2.824
0.72	0.007	0.003	0.004	664	-2.423	2.822
0.73	0.007	0.003	0.004	666	-2.413	2.823
0.85	0.009	0.003	0.005	674	-2.299	2.828
0.84	0.008	0.004	0.005	686	-2.313	2.836
0.83	0.008	0.003	0.005	668	-2.314	2.825
0.96	0.010	0.003	0.006	669	-2.211	2.825
0.95	0.010	0.003	0.006	671	-2.219	2.827
0.96	0.010	0.004	0.006	691	-2.219	2.840
1.07	0.011	0.003	0.007	658	-2.136	2.818
1.07	0.011	0.004	0.007	696	-2.148	2.842

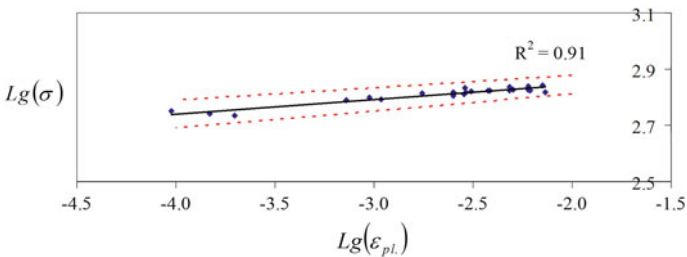


Fig. 11.18 Linear regression for the determination of the cyclic curve parameters, with related lower and upper bounds

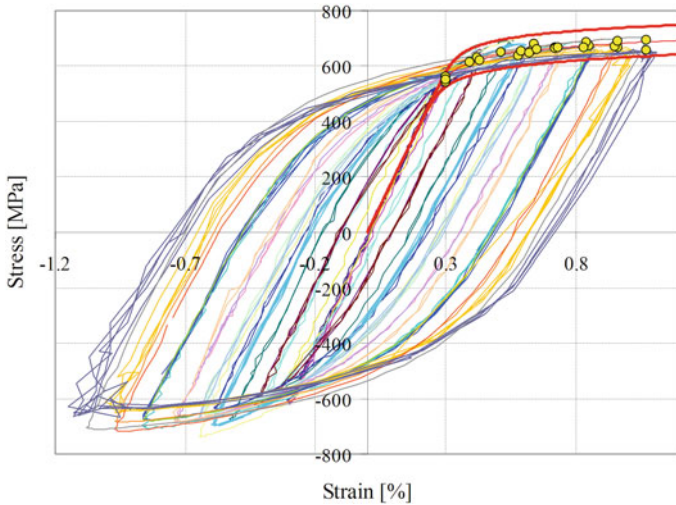


Fig. 11.19 Experimentally determined cyclic curve and bounds wrapped around its plastic part

considering the same scenario of twice the standard deviation, affecting the $\log(K')$ and n' parameters. This scenario can usually be regarded as the worst one, as in [11, 12, 25, 26, 43, 44].

11.3.2 How to Determine the Manson-Coffin Curve Parameters and Related Distributions

The *Manson-Coffin* curve of a material, representative of its low cycle fatigue response must be determined, running a sufficiently high number of fatigue tests in strain controlled conditions. The reference standards are [24, 41, 42], where details on the experimental procedure and on the data processing techniques are provided. The *Manson-Coffin* model is recalled in Eq. 11.20, where σ'_f and ϵ'_f stand respectively for the *fatigue strength* and the *fatigue ductility coefficients*, whereas b and c are the related *fatigue strength* and the *fatigue ductility exponents*.

$$\frac{\Delta\epsilon}{2} = \frac{\Delta\epsilon_{el.}}{2} + \frac{\Delta\epsilon_{pl.}}{2} = \frac{\sigma'_f}{E} \cdot (2N_f)^b + \epsilon'_f \cdot (2N_f)^c \tag{11.20}$$

In Eq. 11.20, the term $\frac{\Delta\epsilon}{2}$ stands for the controlled strain amplitude, while the subscripts $_{el.}$ and $_{pl.}$ are appended to indicate respectively the elastic and plastic amounts of the total strain amplitude. Finally, the symbol N_f refers to the number of cycles to failure and $2N_f$ has the physical meaning of the number of reversals in the stress-strain hysteresis loops.

The parameters $\sigma'_f, \varepsilon'_f, b, c$ are easily determined, separating the amounts of elastic and plastic strain and by running linear regressions in the logarithmic scale. The effect of the application of a 10-base logarithm to the elastic part of the strain amplitude is shown in Eq. 11.21, whereas the same procedure is performed in Eq. 11.22 for the plastic field.

$$\log\left(\frac{\Delta\varepsilon_{el.}}{2}\right) = b \cdot \log(2N_f) + \log\left(\frac{\sigma'_f}{E}\right) \quad (11.21)$$

$$\log\left(\frac{\Delta\varepsilon_{pl.}}{2}\right) = c \cdot \log(2N_f) + \log(\varepsilon'_f) \quad (11.22)$$

Therefore, the *fatigue strength* and the *fatigue ductility exponents* can be computed as the slopes of the regression lines, whereas the corresponding *coefficients* are retrieved as regression line offset terms.

As for the static/cyclic curve, in this case it is also important to make a literature survey, regarding the statistical distributions to be presumed for the parameters involved in the *Manson-Coffin* model. The *fatigue strength* and *ductility coefficients*, σ'_f and ε'_f , are usually regarded as log-normally distributed random variables. This approach is confirmed by [36, 37] and by more recent studies [11, 12, 32, 33], even if in one case [38] normal distributions were considered. Regarding the *fatigue exponents*, b and c , some studies [36, 37] accounted for them as deterministic constants, whereas in all the other cases [11, 12, 32, 33, 38], they were supposed to have normal distributions.

Therefore, according to the most widely applied approach, σ'_f and ε'_f are log-normally distributed, which means that their logarithms $\log(\sigma'_f)$ and $\log(\varepsilon'_f)$ follow normal distributions. The *exponents* b and c are also normally distributed. Therefore, the determination of the standard deviations of the slopes and of the offset terms of the regression line, as in Sect. 11.4, makes it possible to univocally determine the normal distributions of the aforementioned terms. It is clear that the standard deviation of $\log\left(\frac{\sigma'_f}{E}\right)$ is the same as that of $\log(\sigma'_f)$, as the Young's modulus E is deterministic.

Exercise 11.2 (*Manson-Coffin curve of a steel material: distributions of the related coefficients and exponents*) A numerical example of the procedure in Sect. 11.3.2 is provided here.

With reference to the same steel material of Sect. 11.1, characterized in agreement with [41, 42], the results, in terms of the controlled strain amplitudes $\left(\frac{\Delta\varepsilon}{2}\right)$, of their elastic $\left(\frac{\Delta\varepsilon_{el.}}{2}\right)$ and plastic $\left(\frac{\Delta\varepsilon_{pl.}}{2}\right)$ parts and of the retrieved cycles to failure (N_f), are listed in Table 11.12 [11, 26].

The model in Eq. 11.21, with reference to the elastic part, leads to a rearrangement of the data, with the computation of the $\log(2N_f)$ and $\log\left(\frac{\Delta\varepsilon_{el.}}{2}\right)$ terms. The reworked yields, to be linearly interpolated, are shown in Table 11.13.

Table 11.12 Results of the low cycle fatigue tests and observed lives N_f

Total strain amplitude $\left(\frac{\Delta\varepsilon}{2}\right)$ (%)	Total strain amplitude $\left(\frac{\Delta\varepsilon}{2}\right)$ (-)	Elastic strain amplitude $\left(\frac{\Delta\varepsilon_{el.}}{2}\right)$ (-)	Plastic strain amplitude $\left(\frac{\Delta\varepsilon_{pl.}}{2}\right)$ (-)	Cycles to failure $(2N_f)$
0.20	0.002	0.002	1.13×10^{-5}	50,258
0.20	0.002	0.002	1.76×10^{-5}	51,516
0.30	0.003	0.003	1.30×10^{-4}	17,158
0.30	0.003	0.003	1.55×10^{-4}	14,575
0.43	0.004	0.003	0.001	4,885
0.43	0.004	0.003	0.001	4,719
0.50	0.005	0.003	0.002	1,080
0.50	0.005	0.003	0.002	1,100
0.58	0.006	0.003	0.002	905
0.58	0.006	0.003	0.002	1,117
0.63	0.006	0.003	0.003	749
0.65	0.007	0.003	0.003	385
0.65	0.007	0.003	0.003	779
0.73	0.007	0.003	0.004	449
0.73	0.007	0.003	0.004	347
0.85	0.009	0.003	0.005	260
0.85	0.009	0.004	0.005	296
0.96	0.010	0.004	0.006	250
0.96	0.010	0.004	0.006	251
1.07	0.011	0.004	0.007	100

The linear regression of the points in the two columns of Table 11.13 leads to the computation of the slope and of the offset term, namely $b = -0.043$ and $\log\left(\frac{\sigma'_f}{E}\right) = -2.340$. According to the procedure described in Sect. 11.4, the following terms can be easily calculated, based on the data in Table 11.13.

- $S = 18$
- $S_x = 58.48$
- $S_y = -44.61$
- $S_{xx} = 196.66$
- $S_{xy} = -145.22$
- $\Delta = 119.70$
- $\bar{\sigma} = 0.0081$

Finally, the previous results make it possible to calculate the standard deviations of the slope b and of the offset $\log\left(\frac{\sigma'_f}{E}\right)$ are respectively 3.1×10^{-3} and $1.03 \times$

Table 11.13 Data for the linear interpolation in the elastic field

$\log(2N_f)$	$\log\left(\frac{\Delta\epsilon_{el.}}{2}\right)$
4.535	-2.542
4.465	-2.546
3.990	-2.494
3.975	-2.496
3.334	-2.484
3.342	-2.484
3.258	-2.482
3.349	-2.475
3.176	-2.471
2.886	-2.469
3.193	-2.475
2.953	-2.464
2.841	-2.464
2.716	-2.458
2.772	-2.456
2.699	-2.451
2.701	-2.452
2.297	-2.448

10^{-2} . Regarding the *fatigue strength coefficient*, its logarithm can be easily obtained, considering that $E = 193.8$ GPa, therefore we have a normal distribution with mean value $\log(\sigma'_f) = 2.948$ and the aforementioned standard deviation of 1.03×10^{-2} . Operating in the same way, considering the model in Eq. 11.22, it is possible to determine the normal distributions of c (mean value: -0.546 , standard deviation: 3.9×10^{-2}) and of $\log(\epsilon'_f)$ (mean value: -0.83 , standard deviation: 0.12).

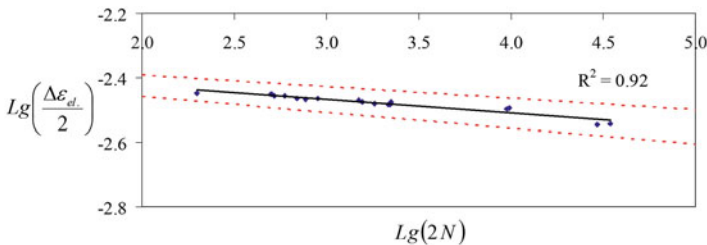


Fig. 11.20 Linear regression for the determination of σ'_f and b , with related *lower* and *upper* bounds

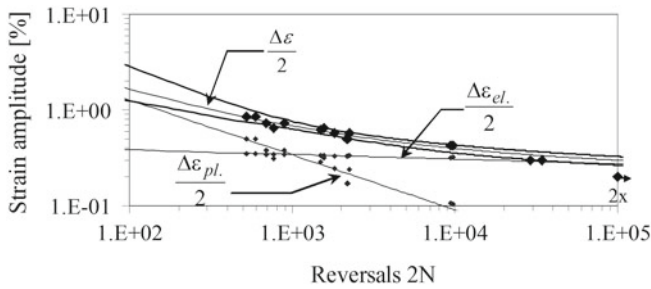


Fig. 11.21 Experimentally determined *Manson-Coffin* curve together with wrapped around hyperbolic bounds

The retrieved interpolating line is shown in the graph in Fig. 11.20 in a $\log\left(\frac{\Delta\epsilon_{el.}}{2}\right) - \log(2N_f)$ reference system; lower and upper bounds for the worst scenario of twice the standard deviation are also appended. Finally, the determined fatigue curve is depicted in Fig. 11.21 with indication of the experimental dots and of the results in the elastic and plastic fields. The curve is plotted along with its hyperbolic bounds, determined according to the recommendations of [24] and to the original procedure proposed in [26].

11.4 Multi-variable Problems: Application of a *Most Probable Point Method*

In many applications the output variable is a function of more than one variable, therefore a relationship in a multi-dimensional space must be used. In addition, many complications often arise from two issues. First of all, the relationships between the inputs and the output are often non-linear. Regarding this, possible examples are the relationships that relate the expected life to the load entity and to material parameters in high- or low-cycle fatigue and in fracture mechanics. The second issue is concerned with the determination of a closed form analytical relationship that is able to directly yield the output as a function of the inputs. Unfortunately, this closed form solution is often unavailable and impossible to retrieve. For instance, in low cycle fatigue the strain amplitude to be used as an input in the application of the strain-life model must be determined by simulating the stress-strain hysteresis loops. This procedure requires the numerical solution of a set of inter-related non-linear equations. These kind of functional relationships are usually indicated as *black-box* or *implicit* functions. Moreover, a numerical approach is usually required to overcome the non-linearities.

Let g indicate the functional relationship between the output and the inputs. A *Most Probable Point (MPP) Method*, e.g. the *Advanced Mean Value (AMV)* [37] utilizes a *Taylor* expansion to achieve a suitable polynomial approximation of the unknown

function g in the neighbourhood of an *MPP*. The *MPP* is usually referenced as the *design point*: it indicates a particular combination of the values of the input variables, which is likely to result in a failure. The physical meaning of the *design point* will be clarified below. As a consequence of the need to achieve a polynomial approximation in the surroundings of the *design point*, this point and related coordinates should be chosen as the *expanding point* for the expansion. However, a further criticality arises from the occurrence that the *design point* is not known *a priori*, consequently an iterative approach is required for its determination. Therefore, the *expanding point* for the *Taylor* expansion is conventionally initially set as the point defined by the mean values of the input variables: $\mathbf{a} = [a_1, \dots, a_r]^T$. A number of r input variables is considered here, denoting the related vector as $\mathbf{U} = [U_1, \dots, U_r]^T$. It is assumed that all the variables are normally distributed. Should some of them be affected by a different distribution, they could be transformed into equivalent normal variables by the *Rosenball* transformation. For full details regarding its application, Ref. [45] is recommended.

Different approaches can be used in the determination of the polynomial expression of g by *Taylor* expansion: first- or second-order models are usually applied. The linear models are the most widely used, according to [11, 12, 33, 37] and may be suitable, even if the output-input relationship is not linear. Whereas, when great non-linearities are present, second order models are required. Both approaches will be described in the following lines.

According to a first-degree *Taylor* expansion, the function g can be approximated by the polynomial Y , as in Eq. 11.23.

$$\begin{aligned}
 Y(U_1, \dots, U_r) &= g(\mathbf{a}) + \frac{\partial g}{\partial U_1} |_{(\mathbf{a})} \cdot (U_1 - a_1) + \dots + \frac{\partial g}{\partial U_i} |_{(\mathbf{a})} \cdot (U_i - a_i) + \dots \\
 &\quad + \frac{\partial g}{\partial U_r} |_{(\mathbf{a})} \cdot (U_r - a_r) + (H.O.T.) \\
 &= g(\mathbf{a}) + \sum_{i=1}^r \left[\frac{\partial g}{\partial U_i} |_{(\mathbf{a})} \cdot (U_i - a_i) \right] \\
 &= \alpha_0 + \alpha_1 \cdot (U_1 - a_1) + \dots + \alpha_i \cdot (U_i - a_i) + \dots + \alpha_r \cdot (U_r - a_r)
 \end{aligned} \tag{11.23}$$

The term *H.O.T.* in the first-order expansion in Eq. 11.23 stands for negligible *higher order terms*. The derivative terms $(\alpha_1, \dots, \alpha_r)$ may be numerically calculated by the finite difference methods. In particular, the partial derivatives are approximated by quotients of finite differences, determined by operating slight perturbations of the random variables in the neighbourhood of the *expanding point*. Equation 11.24 shows the computation of the (i th) derivative term, by the forward finite difference approach [2], considering a positive increment affecting the (i th) variable.

$$\frac{\partial g}{\partial U_i} |_{(\mathbf{a})} \approx \frac{\Delta g}{\Delta U_i} |_{(\mathbf{a})} = \frac{g(a_1, \dots, a_i + \Delta U_i, \dots, a_r) - g(a_1, \dots, a_i, \dots, a_r)}{\Delta U_i} \tag{11.24}$$

The term ΔU_i in Eq. 11.24 stands for the entity of the (*i*th) perturbation: it must be a good compromise from being too high (which would imply a rough estimation of the partial derivative punctual value) and being too low (which would imply a very stiff polynomial approximation, being reliable only in a small surrounding of the *expanding point*). For this purpose, Refs. [36, 37] suggest choosing the perturbation size ΔU_i as 10% of the standard deviation of the (*i*th) variable. Therefore a suitable choice for the increment is $\Delta U_i = 0.1 \cdot STD(U_i)$, where *STD* indicates the standard deviation. It can be pointed out that a linear model requires in general $(r + 1)$ evaluations for the computation of the α_i derivative coefficients in the last row of Eq. 11.23 [37].

If we skip to a non-linear model, a polynomial approximation of the relationship *g* is yielded by a second-order *Taylor* expansion, as in Eq. 11.25.

$$\begin{aligned}
 & Y(U_1, \dots, U_r) \\
 &= g(\mathbf{a}) + \frac{\partial g}{\partial U_1} |_{(\mathbf{a})} \cdot (U_1 - a_1) + \dots + \frac{\partial g}{\partial U_i} |_{(\mathbf{a})} \cdot (U_i - a_i) \\
 &+ \dots + \frac{\partial g}{\partial U_r} |_{(\mathbf{a})} \cdot (U_r - a_r) + \frac{1}{2} \frac{\partial^2 g}{\partial U_1^2} |_{(\mathbf{a})} \cdot (U_1 - a_1)^2 \\
 &+ \dots + \frac{1}{2} \frac{\partial^2 g}{\partial U_i^2} |_{(\mathbf{a})} \cdot (U_i - a_i)^2 + \dots + \frac{1}{2} \frac{\partial^2 g}{\partial U_r^2} |_{(\mathbf{a})} \cdot (U_r - a_r)^2 \\
 &+ 2 \sum_{i=1}^{r-1} \left\{ \sum_{j=i+1}^r \left[\frac{\partial^2 g}{\partial U_i \partial U_j} |_{(\mathbf{a})} \cdot (U_i - a_i) (U_j - a_j) \right] \right\} + (H.O.T.) \\
 &= g(\mathbf{a}) + \sum_{i=1}^r \left[\frac{\partial g}{\partial U_i} |_{(\mathbf{a})} \cdot (U_i - a_i) \right] + \frac{1}{2} \sum_{i=1}^r \left[\frac{\partial^2 g}{\partial U_i^2} |_{(\mathbf{a})} \cdot (U_i - a_i)^2 \right] \\
 &+ 2 \sum_{i=1}^{r-1} \left\{ \sum_{j=i+1}^r \left[\frac{\partial^2 g}{\partial U_i \partial U_j} |_{(\mathbf{a})} \cdot (U_i - a_i) (U_j - a_j) \right] \right\} + (H.O.T.) \\
 &\cong \alpha_0 + \alpha_1 \cdot (U_1 - a_1) + \dots + \alpha_i \cdot (U_i - a_i) + \dots + \alpha_r \cdot (U_r - a_r) \\
 &+ \alpha_{r+1} \cdot (U_1 - a_1)^2 + \dots + \alpha_{r+i} \cdot (U_i - a_i)^2 + \dots + \alpha_{2r} \cdot (U_r - a_r)^2 \\
 &= \alpha_0 + \sum_{i=1}^r [\alpha_i \cdot (U_i - a_i)] + \sum_{i=1}^r [\alpha_{r+1} \cdot (U_i - a_i)^2] \tag{11.25}
 \end{aligned}$$

It can be observed that the polynomial expression in its compact form consists of four terms: the first and the second are the same as the first-order model, whereas the third and the fourth ones contain higher order derivatives. In particular, the last one, containing mixed derivatives, is usually regarded as the *mixed term*. In Ref. [37] it is remarked that the number of evaluations required for the computation of all the derivative terms in Eq. 11.25 is $\frac{(r+2) \cdot (r+1)}{2}$. The size of the data set is more conveniently reduced to $(2r + 1)$, when the mixed term is neglected. The question related to whether the *mixed term* should be considered or discarded is discussed in some Refs., eg. [12, 37]. The conclusion is that neglecting the *mixed term* not only simplifies the computational procedure, but has also an additional positive outcome.

It derives from the occurrence that in the polynomial approximation of highly non-linear functions, problems of numerical ill-conditioning may arise from this term, thus reducing the accuracy of the *Taylor* expansion.

Considering the first three terms, the unknown coefficients $\alpha_0, \alpha_1, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_{2r}$ can be determined by $(2r + 1)$ function evaluations. According to [12, 37], the data sets can be conveniently determined as in Table 11.14, where the ΔU_i terms retain the same meaning as slight perturbations of the random variables around their values at the *expanding point*. Considering the data set along each row, the function output value Y , indicated at the right side of the same line, must be computed. Subscripts from 1 to $(2r + 1)$ are appended to relate the output values to the corresponding data set.

Following the plan for function evaluations in Table 11.14 makes it possible to compute the $(2r + 1)$ α_i values as unknowns of the linear system in Eq. 11.26.

$$\begin{aligned}
 \alpha_0 &= Y_0 \\
 \alpha_0 + \alpha_1 \cdot \Delta U_1 + \alpha_{r+1} \cdot (\Delta U_1)^2 &= Y_1 \\
 \alpha_0 - \alpha_1 \cdot \Delta U_1 + \alpha_{r+1} \cdot (-\Delta U_1)^2 &= Y_2 \\
 \dots & \\
 \alpha_0 + \alpha_i \cdot \Delta U_i + \alpha_{r+i} \cdot (\Delta U_i)^2 &= Y_{2i-1} \\
 \alpha_0 - \alpha_i \cdot \Delta U_i + \alpha_{r+i} \cdot (-\Delta U_i)^2 &= Y_{2i} \\
 \dots & \\
 \alpha_0 + \alpha_r \cdot \Delta U_r + \alpha_{2r} \cdot (\Delta U_r)^2 &= Y_{2r-1} \\
 \alpha_0 - \alpha_r \cdot \Delta U_r + \alpha_{2r} \cdot (-\Delta U_r)^2 &= Y_{2r}
 \end{aligned}
 \tag{11.26}$$

The rest of the numerical procedure is the same for both linear and second-order models. The following step consists in the formulation of a *failure function*. This passage requires a brief explanation.

The goal of the probabilistic analysis consists in the determination of the probability of $Y \leq y$, where y is a threshold value for Y . The curve relating the probability to y , for different threshold values, is called the *cumulative distribution function (cdf)*

Table 11.14 Data sets for the computation of α_i constant terms

U_1	...	U_i	...	U_r	Y
a_1	...	a_i	...	a_r	Y_1
$a_1 + \Delta U_1$...	a_i	...	a_r	Y_2
$a_1 - \Delta U_1$...	a_i	...	a_r	Y_3
...
a_1	...	$a_i + \Delta U_i$...	a_r	Y_{2i-1}
a_1	...	$a_i - \Delta U_i$...	a_r	Y_{2i}
...
a_1	...	a_i	...	$a_r + \Delta U_r$	Y_{2r-1}
a_1	...	a_i	...	$a_r - \Delta U_r$	Y_{2r}

of Y . The described problem could obviously be tackled by a Monte Carlo simulation, for which full details can be found in specialistic references, e.g. see Birolini [3]. However, this approach often has the drawback of being computationally expensive, especially in the applications with a very high (close to 1) reliability and a very low probability of failure [2].

If we suppose that Y has the meaning of the expected (computed) life for a component under fatigue and that y is the threshold, corresponding to the required life (design specifics), we can have three different cases. If $Y < y$, it means that the expected life is lower than what is required by the specifics, therefore a failure must be expected during the component's life. Consequently, the probability of $Y < y$ assumes the meaning of the *probability of failure* of the studied part. Conversely, $Y > y$ indicates that the expected life is longer than specifics, thus the part should theoretically complete its full life without failures. Finally, the condition of Y being equal to y corresponds to the *boundary* between the unsafe (failure likely to occur) and safe (no failure expected) domains. The function describing this *boundary* is usually regarded as the *failure function* and is easily obtained, subtracting y from Y . In this case, instead of Y , its first- or second-order polynomial approximation is used. It is easy to understand that a *failure function* retains positive values in the safe domain, negative ones in the unsafe region and has zeroes at the boundary. The *failure function* is indicated by h in Eq. 11.27, considering the general case of second-order approximating function. The equation of the boundary between the safe and unsafe domains is therefore expressed by Eq. 11.28.

$$\begin{aligned} h(U_1, \dots, U_i, \dots, U_r) &= Y(U_1, \dots, U_i, \dots, U_r) - y \\ &= \alpha_0 + \sum_{i=1}^r [\alpha_i \cdot (U_i - a_i)] + \sum_{i=1}^r [\alpha_{r+1} \cdot (U_i - a_i)^2] - y \end{aligned} \quad (11.27)$$

$$\alpha_0 + \sum_{i=1}^r [\alpha_i \cdot (U_i - a_i)] + \sum_{i=1}^r [\alpha_{r+1} \cdot (U_i - a_i)^2] - y = 0 \quad (11.28)$$

The further step consists in variable reduction, i.e.: the transformation of the basic variables into the reduced ones. The reduced variables exhibit a *normal* distribution with a mean of zero and standard deviation normalized to one. This procedure is generally performed by the application of Eq. 11.29, which refers to the reduction of the (i th) random variable U_i into the corresponding u_i .

$$u_i = \frac{U_i - \mu(U_i)}{STD(U_i)} \Leftrightarrow U_i = u_i \cdot STD(U_i) + \mu(U_i) \quad (11.29)$$

All the basic variables in the formula of the *failure function*, Eq. 11.28, must therefore be turned into the corresponding reduced variables. For this purpose, the term U_i must be expressed as a function of the term u_i (as in the second member of Eq. 11.29), to be substituted in Eq. 11.28. This computational procedure may be regarded as a

change of the reference system: in the new one the point defined by the mean values of all the input variables is moved to the origin.

Finally, the last calculation step leads to the determination of the *design point* and of its distance from the origin of the new reference system in the reduced variables: this distance is called *safety index*, β . From the analytical point of view, the *design point* and the *safety index* are yielded by the solution of the following constrained optimization problem, Eq. 11.30.

$$\begin{aligned} h(\mathbf{u}^*) &= 0 \\ \beta &= \min \|\mathbf{u}^*\| \end{aligned} \quad (11.30)$$

The symbol * in Eq. 11.30 indicates the reduced coordinates at the *design point*. The formulation of the constrained system of Eq. 11.30 may be explained as follows. The *design point* is unequivocally determined as the point lying on the *failure function* (limit state between the safe and unsafe conditions), being at the minimum distance from the origin of the reduced reference system. It is possible to give a brief explanation of the physical meaning of the *design point*. The *failure function* is the location of all the points, corresponding to different combinations of the values of the input variables, which are likely to lead to a failure. If we suppose that the output variable is fatigue life, then, the points on the curve correspond to states, where the computed expected life equals the life required by design specifics. Among the infinite points with this property, there is a specific point, the *design point*, that has the peculiarity of being the closest to the origin. It corresponds to the input parameter combination that leads to failure and is the closest to the condition, corresponding to all the variables at their mean value. This state (graphically related to the origin) corresponds to all the variables assuming those that are usually reported as their nominal values. The described graphical interpretation, with indication of the *failure function*, *safe* and *unsafe* domains, the *design point* and the *safety index*, is shown in Fig. 11.22. For the sake of clarity, the figure is sketched in a bi-dimensional reference system $(u_i - u_j)$, while, in reality, the *failure function* h should be plotted in an r -dimensional space. From the computational point of view, the constrained system can be easily solved by the *Lagrange multipliers method*, which is summarized in Eqs. 11.31 and 11.32.

$$\lambda_i = \frac{\frac{\partial h}{\partial u_i} \Big|_{\frac{a_i - \mu(U_i)}{STD(U_i)}}}{\sqrt{\sum_{j=1}^r \left[\frac{\partial h}{\partial u_j} \Big|_{\frac{a_i - \mu(U_i)}{STD(U_i)}} \right]^2}} \quad (11.31)$$

$$\mathbf{u}^* = \beta \cdot [\lambda_1, \dots, \lambda_i, \dots, \lambda_r]^T \quad (11.32)$$

In Eq. 11.31, the partial derivatives of the failure function h are calculated at the *expanding point*, expressed here in its reduced coordinates. The application of Eq. 11.31 enables the computation of the terms indicated by λ_i , which are the

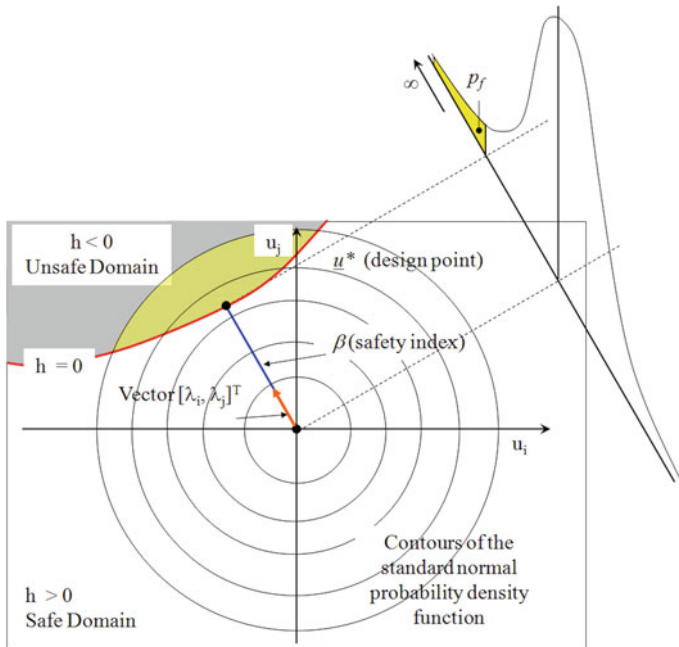


Fig. 11.22 Graphical interpretation of the failure function, separating the same and unsafe domains and of the reliability assessment

direction cosines of a vector that defines the *design point*. Therefore, considering the relationship between \mathbf{u}^* and the $[\lambda_1, \dots, \lambda_i, \dots, \lambda_r]$ vector in Eq. 11.32, the solution of the first equation of Eq. 11.30, yields the *safety index* β . The reduced coordinates of the *design point* \mathbf{u}^* are finally determined by Eq. 11.32, considering the actual value of β .

The probability of failure p_f is finally calculated by Eq. 11.33 [37], where the symbol ϕ stands for the *Standard Normal cumulative distribution function*, with the assumption that the output variable Y is normally distributed [2, 37]. From the graphical point of view, p_f is given by the area of the distribution tail, highlighted in Fig. 11.22.

$$p_f = \phi(-\beta) \tag{11.33}$$

The determined solution is usually regarded as the *Mean value (MV) Solution* [46] and is usually inaccurate. This occurrence can be explained by observing that the result depends on the determined coordinates of the *design point*, determined by Eqs. 11.30, 11.31 and 11.32. All the computation process is based on the first- or second-order polynomial expression (Eqs. 11.23 and 11.25 respectively), relating the output variable to the inputs. As previously stated, this expression is a *Taylor expansion* in the neighbourhood of the *expanding point*. In theory, the expansion should be performed at the *design point* or in the surroundings of it. However, as

remarked at the beginning of this Section, the *design point* is not initially known, therefore, the first *expanding point* is arbitrarily chosen and may even be far away from the real *design point*. This expansion at an arbitrary point is the main reason for the determined polynomial expansion often being a rough approximation of the output-input relationship at the *design point*. As a consequence, the found solution is inaccurate. This problem can be overcome by an iterative procedure. A new *expanding point* is therefore set at the just determined *design point* and a different Taylor expansion is determined. The same algorithm is then repeated for the determination of a new *design point*. At the following step, a new *expanding point* is set at the just determined *design point*, ... Iterations are repeated until convergence. Only a few are usually required: just two or three are usually sufficient [11, 12, 37]. Upon convergence, the final value for the probability of failure p_f can be re-computed by Eq. 11.33.

In order to determine the whole *CDF*, different values of the threshold y must be chosen over a sufficiently wide range. If the output variable Y represents fatigue life, this procedure leads to the determination of a curve that shows the increase of the probability of failure versus life.

11.4.1 Low Cycle Fatigue on Turbogenerators and Probability of Failure

In Sect. 11.4, low cycle fatigue was mentioned as a field, where the discussed *MPP* Method can be applied for a reliability assessment. First of all, it can be remarked that the output, consisting in the observed life, is dependent on many variables. It is indeed dependent on the local state of load, in terms of the strain amplitude, and it is also dependent on the material response under a cyclic load. A further remark regards the non-linearity. Low cycle fatigue occurs, when a particularly severe load is applied, so that the yield strength of the material is locally overcome and plastic deformations occur. It is clear that both the static and the cyclic curves are no longer linear in the plastic range, independently if the *Ramberg-Osgood* model (like in Sect. 11.3.1 and Exercise 11.1) or another one is used for the analytical description. The *Manson-Coffin* strain-life model (used in Sect. 11.3.2 and Exercise 1.2) to be applied for life estimation is also non-linear. A final issue is that they cannot be written in a closed form, i.e. it is not possible to work out a relationship that yields the life as a function of the material parameters and of the state of load. Therefore, the only possible option is trying to determine a polynomial approximation of the relationship between the inputs and the output. A typical application of these concepts is in the design of turbogenerators, whose architecture is shown in Fig. 11.23.

The main components of a turbogenerator are the rotor and the coil retaining rings (CRRs). The machine architecture must be briefly introduced, in particular to clarify the role of the CRRs. The rotor exhibits uniformly spaced longitudinal slots, where copper conductors and related insulating materials are packed. Copper emerges at

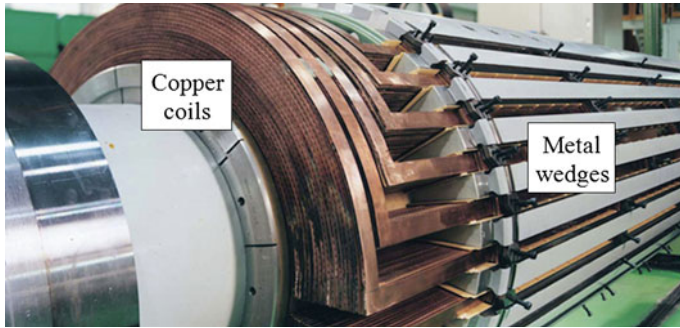


Fig. 11.23 Typical architecture of a turbogenerator

the ends, to join the circumferential arc portion of the windings to form a coil, which is wound around the rotor. During rotation the copper masses must be restrained against the centrifugal forces. This restraint is generally provided by metal wedges applied along the slots. At the ends, where the coils emerge at the surface, wedges cannot be applied and constraining is achieved by the assembly of CRRs that are shrunk fitted onto the rotor body over the coils. The interference coupling generates tangential and radial stresses, acting both on the rotor and on the CRRs.

Turbogenerators are designed for energy production at a constant rotational speed in the order of 3,000 rpm; therefore, the centrifugal force and the induced stresses and strains maintain a constant value during rotation in steady-state conditions, Fig. 11.24. When the machine is switched off, its speed is gradually lowered, and the centrifugal force drops to zero. When switched on, the rotor starts to rotate and its speed is increased to 3,000 rpm. As a consequence, local stresses and strains experience fluctuations at every on and off. This cyclic variation can be regarded as low cycle fatigue, since the expected number of cycles in the entire machine life is about 10,000

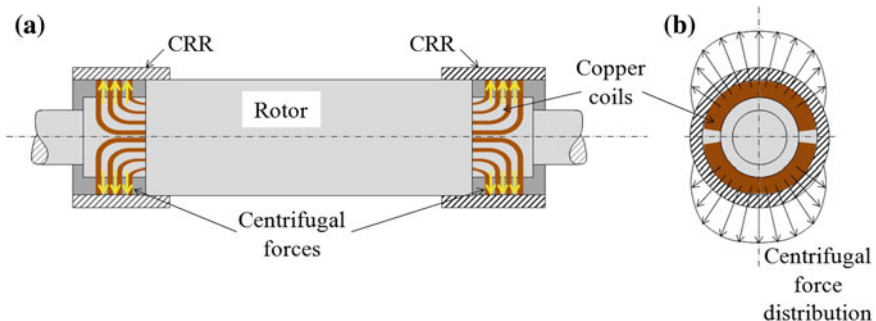


Fig. 11.24 A sketch of the forces acting on the rotor and on the CRR, **a** prospect and **b** lateral views

and has an increasing trend towards 15,000; moreover, the local stresses and strains may exceed the yield limit.

Considering the regions of the rotor and of the CRR at the shrink fit, the stress variation can be qualitatively explained as in the following points.

- Compression of the rotor upon shrink-fit during machine assembly. Compression stresses act both along the radial and the tangential directions. Tensile stresses are generated in the tangential direction on the CRR.
- Machine speed increases up to 3,000 rpm. During rotation, the centrifugal force tends to eject the copper coils, which are restrained by the CRR: as a consequence the CRR swells in tension. Therefore, considering the tangential direction (the direction where the highest stresses occur in both components) the compressive stress is released on the rotor, whereas the tensile stress is incremented on the CRR.
- The speed is lowered down to zero upon machine switch off. The cycle is completed: the compressive stress on the rotor is incremented again, whereas the tensile stress on the CRR is decreased.

Rotor and CRR design are usually tackled by the deterministic application of the models for the simulation of the local stress-strain hysteresis loops and for life estimation, based on the estimated local strain amplitudes. However, nowadays there is increasing interest in a probabilistic approach to turbogenerator design, mainly due to two issues. First of all, the consequences of an in service failure involving the rotor or the CRR may be very serious, since an explosion and resultant fire are likely to occur [47–49]. Secondly, the turbogenerator components have a life in the order of about fifty years and the number of switch on and switch off cycles is expected to increase, due to the new policies for energy production. An important question in the last stages of the life is related to the residual life and to the level of safety if the machine is left in service. This question can be tackled only from the probabilistic point of view, considering the variability of the involved random parameters. The determined probability of failure can thus be compared to thresholds mentioned in the literature, mainly in the nuclear and aeronautical fields.

The next question to be tackled is related to the random parameters to be considered. The existing references [11, 12, 32–38] suggest considering the material parameters, in particular the coefficients of the static and cyclic and fatigue curves. Examples dealing with their determination have been shown in Sect. 11.3.1 and Exercise 1.1, Sect. 11.3.2 and Exercise 1.2. The Young's modulus is usually presumed to be deterministic [11, 12, 33, 36–38]. Regarding load, the random properties of the shrink-fit coupling process suggest randomizing the state of load. However, the rotor and the CRRs are coupled under very strict tolerances. Moreover, a severe quality control is implemented, also to prevent any damage to the coils, upon CRR heating during shrink-fit. As far as the rotational speed is concerned, it is accurately controlled and maintained at the set constant value. For this reason, it can be reasonable to regard the load range during each cycle as a deterministic constant. This approximated approach is supported by [11, 12, 50]. In addition, the study performed in

[51] confirmed that if the variability of load is also considered, this has a negligible impact on the reliability assessment.

11.4.2 Case Study: Probability of Failure of a Turbogenerator Rotor

A turbogenerator rotor, made of *26NiCrMoVa 10 5*, undergoes a cyclic load during its life, due to switch ons and switch offs. Its state of load is regarded as deterministic and can be summarized by the hysteresis loops of the local stresses and strains at the shrink-fit location, highlighted in Fig. 11.25. The loops, sketched in the stress-strain diagram in Fig. 11.26, refer to the following three stages:

- initial assembly of the CRR to the rotor;
- pretrial condition during the machine set-up, where the rotor starts rotating up to the speed of 3,600 rpm, before being stopped;
- normal use conditions with frequent switch ons and shut downs and a controlled speed of 3,000 rpm.

The material parameters ($K, n, K', n', \sigma'_f, b, \varepsilon'_f, c$) are regarded as random variables, except for the Young's modulus, considered as a deterministic constant. A summary of the input variables (indicated as U_i) and of the related mean values and standard deviations, identifying their normal distributions, is provided in Table 11.15.

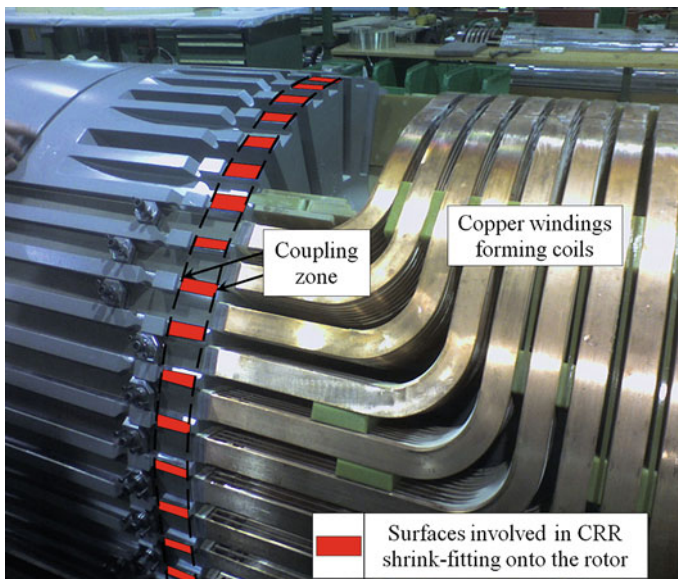


Fig. 11.25 Surfaces of the rotor involved in the shrink-fit

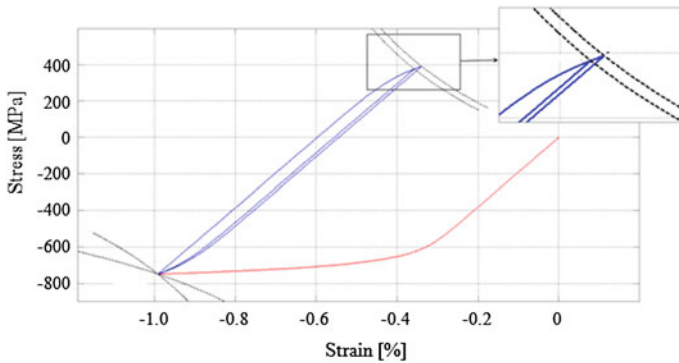


Fig. 11.26 Stress-strain hysteresis loops

Table 11.15 Input random variables and related distributions

Variable	Parameter	Mean value	Standard deviation
U_1	n	0.057	2×10^{-3}
U_2	$\log(K)$	3.001	4×10^{-3}
U_3	n'	0.052	4×10^{-3}
U_4	$\log(K')$	2.949	10^{-2}
U_5	$\log(\sigma'_f)$	2.948	10^{-2}
U_6	$\log(\varepsilon'_f)$	-0.830	0.122
U_7	b	-0.043	3×10^{-3}
U_8	c	-0.546	3.9×10^{-2}

The reliability assessment, with the computation of the safety index and of the failure probability of the rotor during its life, can be implemented, following the procedure described in Sect. 11.4 (see also [11]).

The results in terms of the *safety index* β versus life are shown in Fig. 11.27. It can be observed that β assumes values higher than 4 in the whole machine life range. A threshold for the *safety index* is usually set at 3 [52]. *Safety index* reference ranges for structures under fatigue are contained in [53, 54]. In the case of buildings or bridges, the suggested interval is between 1.5 and 3.8, while in ship design the target *safety index* is usually between 2 and 3.5.

The *cdf* of the probability of failure versus the machine life is plotted in Fig. 11.28. The results are compared here with those of a Monte Carlo simulation with 5×10^6 iterations. The outcomes can be commented on the basis of two issues: efficiency and accuracy.

Regarding the first item, a Monte Carlo simulation may require several hours and is unsuitable to estimate very low probability of failure. In this case, it was capable

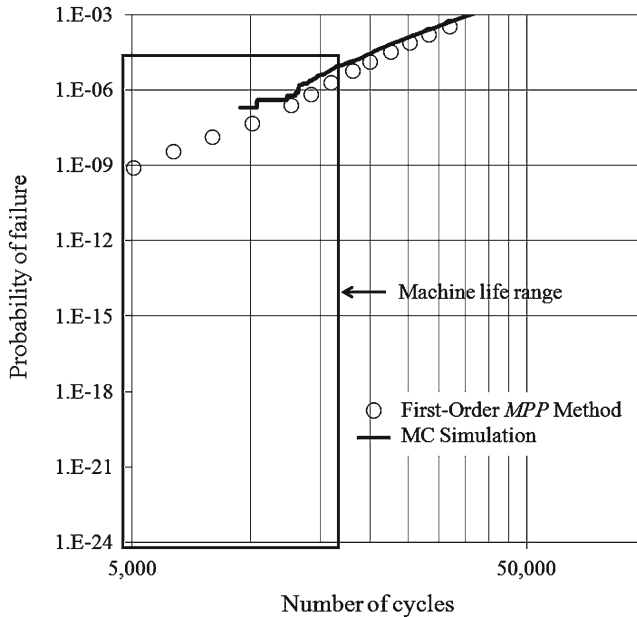


Fig. 11.27 Safety index β plotted versus the machine life

of processing probabilities greater than 10^{-6} , providing just a rough approximation in the range $2 \times 10^{-7} - 10^{-6}$, but was completely unable to provide probability estimates below 2×10^{-7} . Conversely, the applied *MPP* method is highly efficient in the dot-by-dot determination of the *cdf*. Two iterations are generally sufficient to achieve convergence, as confirmed also by [36, 37], which makes it possible to retrieve the *cdf* in just a few minutes. Regarding the second issue, accuracy, the *MPP* method yields results that are comparable to those of a Monte Carlo simulation.

The estimated probability of failure is in the order of 10^{-10} after 6,000 loops, and of 10^{-9} at 8,000 loops and increases up to 5×10^{-8} after 10,000 loops and to 3×10^{-6} at the maximum extended life of 15,000 cycles. These values can be compared with the probabilities of failure of other components, whose failure is potentially highly dangerous, due to its catastrophic effects. An interesting study is reported in [38], regarding the reliability assessment of an aeronautical engine turbine disc structure: in this case the probability of failure is in the order of 10^{-8} .

Therefore, the obtained results would support a turbogenerator life extension up to its maximum of 15,000 cycles, beyond which, generally after a fifty-year life, the machine obsolescence becomes significant.

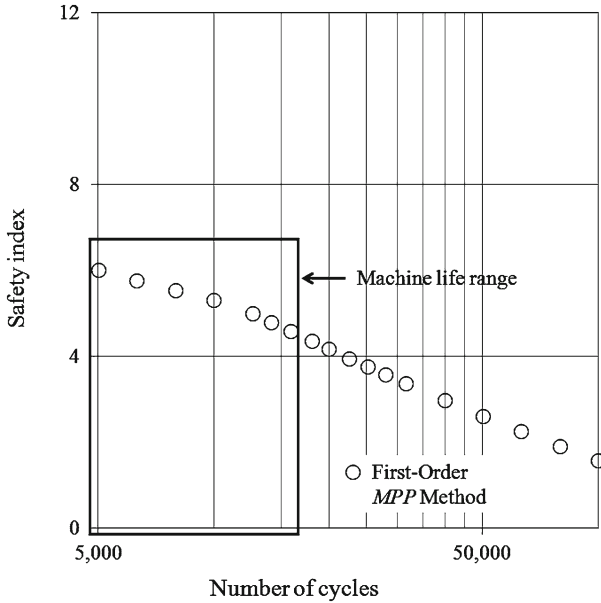


Fig. 11.28 Probability of failure p_f plotted versus the machine life

11.4.3 Fracture Mechanics and Reliability Assessment

A common problem in the mechanics of materials is related to the generation of cracks. Under a cyclic load cracks tend to propagate, until a critical length is reached. When it takes place, the propagation mode turns from stable to unstable and complete breakage is likely to occur. Being able to predict the life of a mechanical part where a crack is present can be very important, when a component is designed according to a *damage tolerant* strategy, an approach that presumes the presence of the crack. In many industrial applications, components undergo periodical refurbishments, where they are checked by non-destructive techniques for the presence of internal or surface cracks. A really difficult question to be tackled is related to the residual life, should cracks be detected. In many cases, for instance in the field of nuclear or power plants, or in aeronautics, in-service failures may even lead to catastrophic events. Therefore, the most proper approach is to tackle the aforementioned question from the probabilistic point of view, determining a probability of failure to be compared to a threshold. For this purpose, a Monte Carlo simulation or an *MPP* method can be applied.

Regarding the tools for life prediction, the most widely used in Fracture Mechanics is the *Paris Law*, reported in Eq. 11.34.

$$\frac{da}{dN_f} = C \cdot \Delta K^r \tag{11.34}$$

The two coefficients C and r in Eq. 11.34 are material-dependent parameters, whereas ΔK stands for the range of the stress intensity factor K . Finally, a is the crack length and N_f stands for the number of cycles to failure, the derivative term $\left(\frac{da}{dN_f}\right)$ has the meaning of the crack propagation rate. This relationship can be integrated to determine life estimation: the final result is shown in Eq. 11.35.

$$N_f = \frac{2}{C \cdot (r - 2) \cdot (M \Delta \sigma \sqrt{\pi})^r} \cdot \left(\frac{1}{a_0^{\frac{r-2}{2}}} - \frac{1}{a_f^{\frac{r-2}{2}}} \right) \quad (11.35)$$

The term $\Delta \sigma$ stands for the stress range far away from the crack, this term is related to the stress intensity factor by the well known relationship. $\Delta K = M \Delta \sigma \sqrt{\pi a}$, where M is a magnifying factor, depending on the crack shape and on the loading model. Finally, a_f stands for the final length of the crack upon breakage. This term can be determined in the case of pulsating load, as in Eq. 11.36, presuming that the crack propagation becomes unstable, when $\Delta K = K_{Ic}$ and when $a_f = a_c$, where the subscript c stands for *critical*. The parameter K_{Ic} is a material property, the well known *fracture toughness*.

$$\Delta K = K_{Ic} \Leftrightarrow M \Delta \sigma \sqrt{\pi a_c} = K_{Ic} \Leftrightarrow a_c = \frac{1}{\pi} \cdot \left(\frac{K_{Ic}}{M \Delta \sigma} \right)^2 \quad (11.36)$$

Considering Eqs. 11.36 and 11.35 can be reworked as in Eq. 11.37.

$$N_f = \frac{2}{C \cdot (r - 2) \cdot (M \Delta \sigma \sqrt{\pi})^r} \cdot \left[\frac{1}{a_0^{\frac{r-2}{2}}} - \frac{1}{\left(\frac{K_{Ic}}{M \Delta \sigma \sqrt{\pi}} \right)^{r-2}} \right] \quad (11.37)$$

The uncertainty in the estimation of the residual life N_f , regarded as the output variable Y , by Eq. 11.37, usually arises from the material response under fracture, in particular from the parameters, C , r and K_{Ic} and from the estimation of the initial length of the crack a_0 . This term may be the source of high errors in the prediction of life, especially when it is low with respect to a_c . In this condition, a small error in the estimation of a_0 may result in a huge error in that of the residual life. Finally, a further source of uncertainty may be the nominal load $\Delta \sigma$, even if, when considering a component, whose cyclic load is strictly controlled, it is reasonably a constant.

11.4.4 Case Study: Fracture Mechanics and Reliability Assessment

A numerical example of the application of an *MPP* method for the probabilistic analysis of the residual life in fracture mechanics is provided here.

A press column is a typical component that withstands a particularly severe pulsating tensile load, with one cycle being completed at every pushing-in and pushing-out. We suppose that a crack has initiated on the column and is experiencing growth under the periodic load. Determining the residual life of the press is really a crucial problem. However, uncertainties may arise from the material fracture behaviour and from the crack actual extension, often difficult to accurately estimate. The distributions of C , r , K_{Ic} and a_0 are reported in Table 11.16. According to [36], normal distributions have been chosen for C , r and K_{Ic} and a log-normal for a_0 . The standard deviations account for the typical uncertainty in the determination of these parameters, also on the basis of [55, 56]. The term $M \Delta\sigma$ is regarded here as a deterministic constant, due to the strict control on the load cycle followed by the machine (see also the remarks at the end of Sect. 11.4.3). Its value, assumed as $M \Delta\sigma = 100$ MPa, can be experimentally determined, for instance by strain gage testing.

The procedure described in Sect. 11.4 can be applied for the reliability assessment. In this case, like for low cycle fatigue (see Sects. 11.4.1 and 11.4.2), the relationship between life and the inputs is again non-linear. However, what can be remarked is that an explicit formula does exist in the present case for the direct computation of life (Eq. 11.35, with the adjustments in Eq. 11.37). The use of a polynomial approximation by *Taylor* expansion makes it possible to obtain a function that can be handled for the computation of the *design point* in an easier way. In order to test the accuracy of the applied *MPP* model, a Monte Carlo simulation with 5×10^5 samplings was also performed. The best matching between results was found here, when considering a second-order model, presumably due to the high non-linearity of Eqs. 11.35 and 11.37. The second-order model also proved to be more efficient, as two or three iterations were generally sufficient to achieve result convergence. The *Taylor* polynomial expansion was determined, running 9 function evaluations, according to the plan of Table 11.14 at p. 478.

The results of the probabilistic study are shown in Fig. 11.29: it can be observed that the probability of failure p_f increases up to about 0.1 % after 30,000 cycles, and then to 1 % after 40,000 and even to 10 % after 60,000 cycles. These are quite high

Table 11.16 Input random variables and related distributions

Variable	Parameter	Mean value	Standard deviation
U_1	C	6.9×10^{-12}	6.9×10^{-13}
U_2	r	3	5×10^{-2}
U_3	K_{Ic}	$60\text{MPa}\sqrt{m}$	$5\text{MPa}\sqrt{m}$
U_4	$\log(a_0)$	-1.347	5×10^{-2}

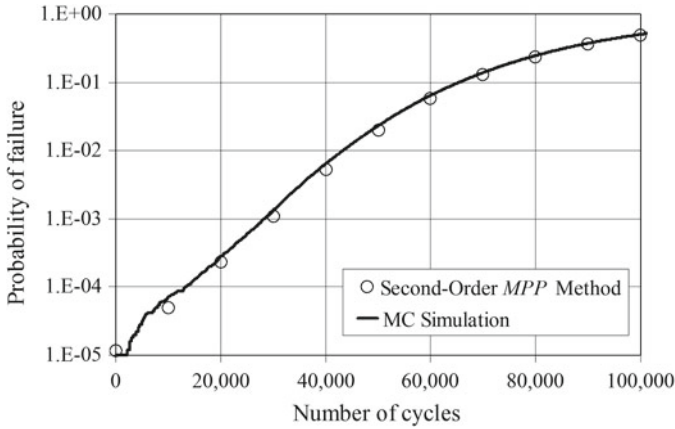


Fig. 11.29 Failure probability p_f prediction, if the press is kept in service

values of the probability of failure, considering that a standing press, involved in industrial production (for instance in the ceramic industry field), may even perform 10,000 cycles or more in just one day.

Therefore, these results indicate that it is highly unsafe to keep the machine working, as the probability of a serious failure is beyond the usually accepted thresholds. This is also confirmed by the trend of the *safety index* β , shown in Fig. 11.30, which drops below the threshold of 3 (see Sect. 11.4.2 and Ref. [52]), after just 40,000 cycles, corresponding to approximately 3 working days.

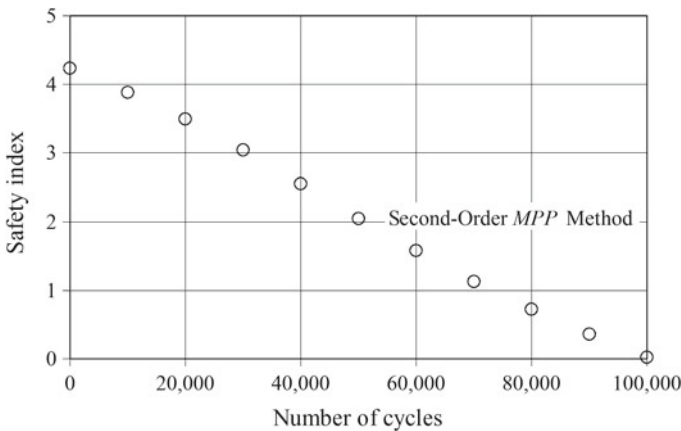


Fig. 11.30 Safety index β prediction, if the press is kept in service

References

1. Helton JC (1997) Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. *J Stat Comput Simul* 57(1–4):3–76
2. Huang B, Du X (2008) Probabilistic uncertainty analysis by mean-value first order saddlepoint approximation. *Reliab Eng Syst Saf* 93(2):325–336
3. Birolini A (2014) *Reliability engineering: theory and practice*, 7th edn. Springer, Berlin
4. Creveling CM, Slutsky JL, Antis D (2003) *Design for six sigma: in technology and product development*. Prentice Hall, New Jersey
5. Du X, Chen W (2000) Towards a better understanding of modeling feasibility robustness in engineering design. *ASME J Mech Des* 122(4):385–394
6. Du X, Chen W (2004) Sequential optimization and reliability assessment for probabilistic design. *ASME J Mech Des* 126(2):225–233
7. Du X, Sudjianto A, Chen W (2004) An integrated framework for optimization under uncertainty using inverse reliability strategy. *ASME J Mech Des* 124(4):562–570
8. Haldar A, Mahadevan S (2001) *Probability, reliability, and statistical methods in engineering design*. Wiley, New York
9. Mailhot A, Villeneuve JP (2003) Mean-value second-order uncertainty analysis method: application to water quality modeling. *Adv Water Resour* 26(5):491–499
10. Nikolaidis E, Chen S, Cudney H, Hatftka RT, Rosca R (2004) Comparison of probability and possibility for design against catastrophic failure under uncertainty. *ASME J Mech Des* 126(3):386–394
11. Olmi G (2012) An efficient method for the determination of the probability of failure on the basis of LCF data: application to turbogenerator design. *SDHM Struct Durab Health Monit* 8(1):61–89
12. Olmi G, Freddi A (2014) Reliability assessment of a turbogenerator coil retaining ring based on low cycle fatigue data. *Arch Mech Eng* 61(1):5–34
13. Putko MM, Newman PA, Taylor A, Green LL (2002) Approach for uncertainty propagation and robust design in cfd using sensitivity derivatives. *ASME J Fluids Eng* 124(1):60–69
14. Seo HS, Kwak BM (2002) Efficient statistical tolerance analysis for general distributions using three-point information. *Int J Prod Res* 40(4):931–944
15. Taguchi G (1993) *Taguchi on robust technology development: bringing quality engineering upstream*. ASME, New York
16. Wu YT, Shin Y, Sues R, Cesare M (2001) Safety-factor based approach for probabilistic-based design optimization. In: *Proceedings of 42nd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and material conference and exhibition*, Seattle, Washington
17. Youn BD, Choi KK (2004) An investigation of nonlinearity of reliability-based design optimization approaches. *ASME J Mech Des* 126(3):403–411
18. Olmi G (2015) Statistical tools applied for the reduction of the defect rate of coffee degassing valves. *Case Stud Eng Fail Anal* 3:17–24. doi:[10.1016/j.csefa.2014.10.002](https://doi.org/10.1016/j.csefa.2014.10.002)
19. Belingardi G (2014) *Strumenti statistici per la meccanica sperimentale e l'affidabilità*. Levrotto & Bella, Turin
20. Keller G (2014) *Statistics for management and economics*, 10th edn. Cengage Learning, Boston
21. Chambers JM, Cleveland WS, Tukey PA, Kleiner B (1983) *Graphical methods for data analysis*. Duxbury Press, Pacific Grove
22. Levine DM, Krehbiel TC, Berenson ML (2010) *Business statistics: a first course*. Prentice Hall, Upper Saddle River
23. EN ISO 6892-1:2009 (2009) *Metallic materials—tensile testing. Part 1: method of test at room temperature*
24. ASTM E739-10 (2010) *Standards practice for statistical analysis of linear or linearized stress-life (S-N) and strain-life (ϵ -N) fatigue data*. ASTM International, West Conshohocken, PA, USA
25. ISO 12107:2012 (2012) *Metallic materials—fatigue testing—statistical planning and analysis of data*. International Organization for Standardization (ISO), Geneva, Switzerland

26. Olmi G (2012b) Low cycle fatigue experiments on turbogenerator steels and a new method for defining confidence bands. *J Test Eval (JTE)* 40(4):539–552
27. Doyle JF (2004) *Modern experimental stress analysis. Completing the solution of partially specified problems*, vol 1. Wiley, New York (ISBN 0-470-86156-8)
28. Gardner L, Nethercot DA (2004) Experiments on stainless steel hollow part 1: material and cross-sectional behaviour. *J Constr Steel Res* 60:1291–1318
29. Hertelé S, De Waele W, Denys R (2011) A generic stress-strain model for metallic materials with two-stage strain hardening behaviour. *Int J Non-Linear Mech* 46:519–531
30. Ramberg W, Osgood WR (1943) Description of stress-strain curves by three parameters. Technical report 902, NACA
31. Rasmussen KJR (2003) Full-range stress-strain curves for stainless steel alloys. *J Constr Steel Res* 59:47–61
32. Choi KK, Youn BD (2002) On probabilistic approaches for reliability-based design optimization. In: *Proceedings of 9th AIAA/NASA/USA/ISSMO symposium on multidisciplinary analysis and optimization*, Atlanta, Georgia
33. Grujicic M, Arakere G, Bell WC, Marvi H, Yalavarthy HV, Pandurangan B, Haque I, Fadel GM (2010) Reliability-based design optimization for durability of ground vehicle suspension system components. *J Mater Eng Perform* 19(3):301–313
34. Zhang ZP, Qiao YJ, Sun Q, Li CW, Li J (2009) Theoretical estimation to the cyclic strength coefficient and the cyclic strain-hardening exponent for metallic materials: preliminary study. *J Mater Eng Perform* 18(3):245–254
35. Zhu SP, Huang HZ, Ontiveros V, He LP, Modarres M (2012) Probabilistic low cycle fatigue life prediction using an energy-based damage parameter and accounting for model uncertainty. *Int J Damage Mech* 21(8):1128–1153
36. Wirsching PH, Torng TY, Martin WS (1991) Advanced fatigue reliability analysis. *Int J Fatigue* 13(5):389–394
37. Wu YT, Wirsching PH (1984) Advanced reliability method for fatigue analysis. *J Eng Mech* 110(4):536–553
38. Liu CL, Lu ZZ, Xu YL, Yue ZF (2005) Reliability analysis for low cycle fatigue life of the aeronautical engine turbine disc structure under random environment. *Mater Sci Eng A* 395(1–2):218–225
39. Olmi G (2011) A new loading-constraining device for mechanical testing with misalignment auto-compensation. *Exp Tech* 35(6):61–70
40. Olmi G (2012) A novel method for strain controlled tests. *Exp Mech* 52(4):379–393
41. ASTM/E606M-12 (2012) Standard practice for strain-controlled fatigue testing. ASTM International, West Conshohocken
42. ISO 12106:2003 (2003) *Metallic materials—fatigue testing—axial-strain-controlled method*. International Organization for Standardization (ISO), Geneva, Switzerland
43. Dixon WJ, Massey FJ Jr (1983) *Introduction to statistical analysis*, vol 1. McGraw-Hill, New York
44. Olmi G, Freddi A (2013) A new method for modelling the support effect under rotating bending fatigue: application to Ti-6Al-4V alloy, with and without shot peening. *Fatigue Fract Eng Mater Struct* 36(10):981–993
45. Rosenblatt M (1952) Remarks on a multivariate transformation. *Ann Math Stat* 23:470–472
46. Wu YT, Millwater HR, Cruse TA (1990) Advanced probabilistic structural analysis method for implicit performance functions. *AIAA J* 28(9):1663–1669
47. Kilpatrick NL, Schneider MI (1987) Update on experience with in-service examination of non-magnetic rings on generator rotors. In: *Proceedings of generator retaining-ring workshop (EPRI-EL-5825)*. Charlotte, North Carolina, pp 103–109
48. Skorchellety VV, Silina EP, Zaytsev VA, Maslov VV, Lubeznova TD (1981) Stress corrosion cracking of turbogenerators rotors retaining rings. *Elektrychiskie stantsii* 12:40–42
49. Speidel MO (1981) Nichtmagnetisierbare stähle für generator-kappenringe, ihr widerstand gegen korrosionermundung. *VGB Kraftwerkstechnik* 61(5):1048–1053

50. Zhao YX (2000) A methodology for strain-based fatigue reliability analysis. *Reliab Eng Syst Saf* 70(2):205–213
51. Bandini L (2012) Valutazione della probabilità di guasto di rotori per turboalternatori, considerando l'aleatorietà del materiale e della sollecitazione. Bachelor degree thesis, University of Bologna
52. Avriithi K, Ayyub BM (2010) A reliability-based approach for low-cycle fatigue design of class 2 and 3 nuclear piping. *J Press Vessel Technol* 132(51):202-1-6
53. ENV 1991-1 (Eurocode 1) (1994) Basis of design and actions on structures—part I: basis of design. CEN (Comite Européen de la Normalisation), Brussels, Belgium
54. Mansour AE, Wirsching PH, White GJ, Ayyub BM (1992) Probability-based ship design: implementation of design guidelines. Technical report SSC 392, NTIS, Washington
55. Several Authors (1976) Rapid inexpensive tests for determining fracture toughness report of the committee on rapid inexpensive tests for determining fracture toughness. Technical report, National Academy of Sciences, Washington
56. Stephens RI, Fatemi A, Stephens RR, Fuchs HO (2000) *Metal fatigue in engineering*, 2nd edn. Wiley, Hoboken (ISBN 978-0-471-51059-8)

Index

A

Advanced Mean Value, 475
Airy, 229, 232, 370
Ajovalasit, 23, 101
Albert, 265
Aleatory, 450
Alfirevic, 230
Ambramson, 178
Apparent strain, 35
Araldite, 147–149, 218, 223, 237, 238
Atzori, 360

B

Band
 Maximum likelihood band, 466
Basquin, 272
Bedzinski, 436
Biharmonic function, 230
Birefringent, 111, 112, 114, 159
Birolini, 479
Borbas, 158
Bounds
 Cyclic curve bounds, 469
 Fatigue curve bounds, 466, 475
 Hyperbolic bounds, 475
Boussinesq, 217
Brewster, 116

C

Calibration
 Calibration matrix, 72, 79, 86
 Compliance matrix, 71, 79, 86
 Coupling coefficients, 80, 87

Cauchy-Riemann, 370
Charpy, 414
Chrome-Nickel, 27
Circular polariscope, 120, 132
Coherence, 167, 170
 length, 168, 178
Colors, 103, 124, 125, 129, 162
 Absorbed, 124
 Transmitted, 124
Compensator, 128, 131
 Babinet, 131
 Wedge, 128
Conjugate, 114
Constantan, 26
Crack, 488
Critical temperature, 146, 148, 150
Cumulative distribution function, 456, 460, 464
Curioni, 415
Cyclic curve, 332
Cyclic stress/strain curve, 330

D

Damage tolerant strategy, 488
Damped Least Squares Solution, 18
Damping factor, 94
Deformed, 167
Design point, 476
Determined
 Even, 263
Deterministic, 450
Digital photoelasticity, 101
Dislocations, 253

Double refraction or birefringence, 109
Doyle, 45, 389, 396

E

Eichlseder, 287
 Electric field, 102
 Ellipse, 176–178, 181
 Epistemic, 450
 Error
 Systematic error, 33
Euler, 133
 Excursion, 452
 Expanding point, 476
 Extensometer
 Axial extensometer, 330
 Optical extensometer, 331
 Transversal extensometer, 330

F

Failure
 Failure function, 479
 Probability of failure, 451, 460, 479, 484, 486
 Fast and slow axes, 115, 120
 Fatigue
 Fatigue ductility coefficient, 471
 Fatigue ductility exponent, 471
 Fatigue strength coefficient, 471
 Fatigue strength exponent, 471
 Figure of merit, 148, 149
 Finite difference approach, 476
 FKM, 285, 287
 Forces identification, 9
 Frozen stress, 146
 Procedure, 150
 Limitation, 151

G

Gage factor, 33
Generalized Cross Validation, 95
Griffith, 369
Guagliano, 311

H

Haigh, 297
Hamilton, 229
 Hardener, 152
 Harmonic vibration, 103
Hertz, 220, 222
 Hysteresis

 Stress-strain hysteresis loops, 469

Histogram

 Category density, 453, 458, 462
 Category frequency, 453, 457, 462
 Category occurrence, 453, 457, 462
 Category, 452, 457, 461

Hologram, 168–171, 173, 176, 177

Hooke, 5, 40, 116, 146

I

Identification, 263
 Ill-conditioning, 71, 94
 Implicit relationship, 451, 475
 Indigo-violet, 103, 162
Inglis, 369
 Input reproducibility matrix, 13
 Integration of equilibrium equations, 141
 Interference
 Black fringes, 123, 125, 129, 174, 175
Irwin, 367
 Isochromatic order, 124, 126, 129, 130, 132, 158
 Isocline, 126, 134, 136, 139, 141, 150, 154, 156, 158, 160
 Isostatic, 27, 136
 Isotropic, 127, 137

J

Joule effect, 29

K

Kirsch, 229

L

L-Curve, 95
Laermann, 6
Lagrange multipliers method, 480
 Lagrangian multipliers, 16
Lamé-Maxwell, 138
 Laser, 168, 172, 178, 184
Lazzarin, 275
 Life
 Life extension, 451, 487
 Residual life, 484, 489
 Light
 Circular Polarized, 112
 Circularly polarized, 106
 Complex amplitude, 103, 108, 170, 175
 Elliptic polarized, 112
 Elliptical polarized, 107

Extinction, 119, 121, 123, 125, 126, 129,
133–135, 174, 175
Full-wave, 111
Half-wave, 111
Intensity, 114, 170, 174
Modulus, 171, 177
Phase, 170, 171, 175
Plane polarized, 104
Polarized, 104
Quarter-wave, 111, 112, 115–117, 120,
132, 134, 135, 159
Vector, 104
Wavefront, 102
Wavelength, 102, 103, 108, 115, 116,
123–125, 134, 162
Linear matrix algebra, 111
Lippman, 178
Liu, 9
Load cell, 64, 71, 94

M

Maggiore, 350
Manson-Coffin model, 471, 482
Manson-Coffin, 348, 349, 354
Maxwell, 101, 280
Michelson, 178
Minimum Length Solution, 16
Model, 269, 271, 274, 278, 284, 285
FKM, 287
Mean stress, 336
Neuber, 333
Ramberg-Osgood, 331
Reggiani, 288
Three tests, 289
Mohr, 123, 139, 251
Mohr's circles, 383
Mold, 152, 153
Mold design, 152
Monochromatic light, 102, 111
Monomer, 152
Most Probable Point, 475
Musculoskeletal, 425

N

Neuber, 246, 247, 346, 354
Newton, 248
Normal probability plot, 454, 459, 463

O

Output reproducibility matrix, 13

P

Paris, 416
Paris law, 488
Phase, 168
Phase angle, 102
Phase difference, 104, 111, 115, 116
Photoelastic coatings, 159
Photoelastic constant, 116, 122, 139, 148,
223
Photoelastic material, 114
Photoelasticity-Holography analogy, 176
Photoengraving, 29
Piezo-resistive effect, 58
Plane polariscope, 118, 121, 134
Plane Polarizer, 108
Plasticity
Coefficient of plasticity, 468
Hardening exponent, 468
Platinum-Tungsten, 27
Poisson, 223, 230
Polarizer axis, 108
Posed
Ill-posed, 4, 8
Well-posed, 4
Press column, 490
Pressure vessels, 224
Principal strain, 116
Principal stress, 115, 116, 119, 121, 122,
126, 127, 130, 134, 154, 155, 160,
161
Separation, 139, 142
Probabilistic, 451
Probability density function, 456, 460, 464
Propagation front, 102
Pseudo-inverse matrix, 19

Q

Quantile
Quantile-quantile plot, 454, 458, 463
Standardized normal quantile, 454, 459,
463

R

Rainflow, 338, 354
Rainflow counting method, 265
Ramberg-Osgood model, 468
Ramberg-Osgood, 244, 246
Rankine, 303
Reference point, 275, 278, 282, 286, 287,
291, 300, 302, 303, 315, 331, 333
Refractive index, 114
Regression line, 465

Regularization, 5
Regularized Least Squares, 91
 Reliability, 319, 321
 Reversal, 334
 Reversibility, 6
Rice, 366
Rosenball transformation, 476
 Rosette
 Strain gage rosette, 38

S

Safety index, 480, 486, 491
Saint Venant, 217, 224, 227, 234, 241, 269, 275
 Secondary effect, 77
 Secondary principal stress, 143, 150
Senarmont, 128
 Sensitive
 Square sensitive area, 60
 octagonal sensitive area, 60
 Shear stresses, 139
Shih, 366
Siebel and Stieler, 284
 Silicon layer, 58
 Simulation model, 265, 349
 Singular, 127
Singular Value Decomposition, 91
 Ski, 87
 Slender bodies, 217
 Slice, 150, 154–156, 158
Smith, 259
 Stiffness identification, 10
 Strain sensitivity, 116, 144, 148, 149
 Stress
 Multi-axial, 331
 Stress sensitivity, 116, 144, 148
 Stress sign, 139
 Sub-slice, 154, 155

T

Tardy, 128
Taylor, 377, 394
Taylor expansion, 475
 Temporary birefringence, 114
 Three-dimensional photoelasticity, 150
 Threshold
 Acceptability threshold, 451, 484, 486, 491
Tikhonov-Miller, 91
Tikhonov-Phillips regularization method, 94
Tikhonov-Phillips, 91
Timoshenko, 217
Tovo, 275
 Transition Temperature, 414
 Transmission function, 170, 171
 Transpose vector, 114
Tresca, 365, 382
 Turbogenerator, 469, 482
 Two phases theory, 146

U

Uncertainties, 450
 Undeformed, 167

V

Vergani, 311
Von Mises, 279, 300, 365, 381

W

Wöhler, 266
Wells, 366
Westergaard, 369
 Wheatstone bridge, 48, 59, 68, 70

Z

Zero-order fringe, 177